

Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients

RADU ALEXANDRU TODOR[†] AND CHRISTOPH SCHWAB[‡]

Seminar for Applied Mathematics, ETH Zürich, 8092 Zürich, Switzerland

[Received on 20 February 2006; accepted on 18 July 2006]

A scalar, elliptic boundary-value problem in divergence form with stochastic diffusion coefficient $a(x, \omega)$ in a bounded domain $D \subset \mathbb{R}^d$ is reformulated as a deterministic, infinite-dimensional, parametric problem by separation of deterministic ($x \in D$) and stochastic ($\omega \in \Omega$) variables in $a(x, \omega)$ via Karhunen–Loève or Legendre expansions of the diffusion coefficient. Deterministic, approximate solvers are obtained by projection of this problem into a product probability space of finite dimension M and sparse discretizations of the resulting M -dimensional parametric problem. Both Galerkin and collocation approximations are considered. Under regularity assumptions on the fluctuation of $a(x, \omega)$ in the deterministic variable x , the convergence rate of the deterministic solution algorithm is analysed in terms of the number N of deterministic problems to be solved as both the *chaos dimension* M and the multi-resolution level of the sparse discretization resp. the polynomial degree of the chaos expansion increase simultaneously.

Keywords: partial differential equations with stochastic coefficients; Karhunen–Loève expansion; polynomial chaos; sparse tensor-product approximation.

1. Introduction

The numerical solution of elliptic partial differential equations (PDEs) with stochastic input data by deterministic methods has been employed in engineering for several decades now (see, e.g. Ghanem & Spanos, 1997, and the references therein). We distinguish two broad classes of approaches to the deterministic numerical solution of elliptic stochastic PDEs—the perturbation approach and the spectral approach.

The perturbation approach is widely used in engineering applications (see Kleiber & Hien, 1992, and references therein). There are several variants, of which the *first-order second moment* (see, e.g. Dettinger & Wilson, 1981) technique became very popular, all based on *Neumann expansion* of the stochastic solution around its mean field (see Keller, 1964), and successive computation of (in general only) low-order terms in this expansion.

The spectral approach is based on the *Wiener/generalized polynomial chaos* (gPC) expansion (see Wiener, 1938; Schoutens, 2000; Xiu & Karniadakis, 2002) of the input random fields and the random solution combined with either Galerkin projection or collocation in the stochastic variables of the input data. The numerical analysis of this approach has started only quite recently (see, e.g. Babuška *et al.*, 2004, and the references therein).

There, *exponential convergence rates* have been proved with respect to the spectral order of the stochastic discretization, at fixed dimension M of the stochastic parameterization. Since, however,

[†]Email: todor@math.ethz.ch

[‡]Corresponding author. Email: schwab@math.ethz.ch

the number of stochastic variables in Wiener's gPC parameterization of random fields is unbounded, gPC-type formulations parameterize stochasticity of the random data and the random solution by a finite number M of so-called *stochastic* variables. It is essential here that the dimension M of the stochastic variables is a discretization parameter for the input and output random fields and can therefore be arbitrarily large.

Hence, exponential convergence with respect to the degree of tensor-product-type polynomial approximations in these variables as shown, e.g. in Babuška *et al.* (2004), does not, in general, imply low computational complexity of the spectral approach.

More precisely, if tensor-product polynomial discretizations are used in the stochastic variables to ensure consistency, the number of stochastic degrees of freedom (and, hence, the number of deterministic boundary-value problems (BVPs) to be solved) must increase at least exponentially with simultaneously increasing polynomial degree and stochastic dimension, even with adaptive and anisotropic polynomial degree selection. As a consequence, exponential convergence in terms of the number of stochastic degrees of freedom and, hence, in the number of deterministic problems to be solved is lost in space dimension larger than 1, and only subalgebraic convergence rate can be shown (see Frauenfelder *et al.*, 2005).

In this paper, we present new regularity estimates and sparse approximation error bounds for spectral Galerkin and collocation approximations of elliptic PDEs with stochastic coefficients. As in other recent works (e.g. Babuška *et al.*, 2004), we consider as a model problem a diffusion process in a random medium occupying a bounded domain $D \subset \mathbb{R}^d$ with Lipschitz boundary $\Gamma = \partial D$. For brevity of exposition, we focus here mainly on the error analysis of semi-discretization in the stochastic variable.

We model data uncertainty by allowing stochastic diffusion coefficients $a(x, \omega)$, where $\omega \in \Omega$ and (Ω, Σ, P) is a suitable probability space given a set Ω of *events*, a sub-sigma algebra Σ of the Borel sets of admissible diffusion coefficients, and a probability measure P on the set of admissible data.

ASSUMPTION 1.1 We consider $a \in L^\infty(D \times \Omega)$ to be strictly positive, with positive lower and upper bounds a_- and a_+ , respectively,

$$a_- \leq a(x, \omega) \leq a_+, \quad \lambda \times P\text{-a.e. } (x, \omega) \in D \times \Omega. \quad (1.1)$$

Here, λ denotes the Lebesgue measure in \mathbb{R}^d .

The stochastic diffusion problem then reads

$$\begin{cases} -\operatorname{div}(a(\cdot, \omega) \nabla u(\cdot, \omega)) = f(\cdot, \omega) & \text{in } D, \\ u(\cdot, \omega) = 0 & \text{on } \partial D, \end{cases} \quad P\text{-a.e. } \omega \in \Omega. \quad (1.2)$$

The coefficient $a(x, \omega)$ as well as the solution $u(x, \omega)$ are *random fields* in $D \subset \mathbb{R}^d$, i.e. jointly measurable functions from $D \times \Omega$ to \mathbb{R} . Whereas the random field $u(x, \omega)$ is a mathematically well-defined object (see Theorem 1.3 below), the task of computing $u(x, \omega)$ is less obvious to realize numerically and is of limited interest in practice. In applications, only certain statistics and moments of $u(x, \omega)$ are of interest, and this is also our goal of computation, which we formulate as follows.

PROBLEM 1.2 Given statistics (with respect to $\omega \in \Omega$) of the stochastic data a , compute statistics of the random solution u to (1.2), such as the mean field,

$$E_u: D \rightarrow \mathbb{R}, \quad E_u(x) := \int_{\Omega} u(x, \omega) dP(\omega), \quad x \in D,$$

two-point correlation (or higher-order moments),

$$C_u: D \times D \rightarrow \mathbb{R}, \quad C_u(x, x') := \int_{\Omega} u(x, \omega) u(x', \omega) dP(\omega), \quad (x, x') \in D \times D,$$

or probabilistic level sets,

$$D_\varepsilon^\alpha := \{x \in D: P(|u(x, \cdot)| > \alpha) < \varepsilon\}.$$

As mentioned above, good performance of the perturbation approach has been demonstrated in practice (at least for small fluctuations, when the perturbation series could be truncated after the first-order terms). The computation of higher-order terms in the perturbation series (needed in the case of relatively large fluctuations) involves numerical approximation of higher-order moments of the random solution. Using standard discretizations, this results in a loss of linear complexity¹.

Using sparse approximation of the higher-order moments of the data, perturbation algorithms of linear complexity have been developed recently (see Todor, 2005). The results in the present work can be viewed as spectral counterparts of those in Todor (2005) on the convergence of the perturbation approach.

The parameterization of uncertainty is one of the key points in the numerical treatment of problems with stochastic data. A Karh  nen–Lo  ve (KL) expansion separating the deterministic and stochastic variables optimally in the mean-square sense (see, e.g. Lo  ve, 1977, 1978) is a standard procedure to transform the original stochastic problem into a parametric deterministic one. The resulting parameterization belongs to a hypercube of dimension M which is itself a discretization parameter.

The parametric problem is then solved using, e.g. a stochastic Galerkin (sG) method (variationally in both the stochastic parameter and the physical variable; note the need for numerical integration schemes in high-dimensional domains) or by collocation and interpolation in the parametric variable. Back substitution finally gives an approximation to the original stochastic problem and postprocessing is required to obtain statistical information on the random solution. Just as in the case of a Monte Carlo (MC) simulation, detailed information on the joint probability densities of the input data is in general needed. Tensor-product discretization or tensor-product collocation grids in the parametric variable result, however, in *superalgebraic complexity* rates (see, e.g. Babu  ka *et al.*, 2004; Frauenfelder *et al.*, 2005; Matthies & Keese, 2005). This is, in principle, due to the unfavourable scaling of the required computational effort with the parameter dimension M .

The main results of this paper (Theorem 4.11, Theorem 4.17 and Theorem 5.7) are based on explicit constructions of finite element (FE) spaces in the parametric (i.e. stochastic) variables, which are not of tensor-product type, and for which we show here optimal convergence rates of the corresponding Galerkin or Collocation approximations of (1.2) to hold (precisely, algebraic order $p + 1$ for the wavelet-FE-based construction, where p denotes the fixed polynomial degree, and superalgebraic for the p -FE-based polynomial chaos in the sense of Wiener, 1938). Note that the rates are expressed in terms of the number N of deterministic problems to be solved for and are independent of the dimension M of the stochastic variables (see, e.g. Theorem 4.17). We emphasize that our numerical analysis gives, in particular, a concrete, explicit selection of basis functions in the chaos expansion to be used in the stochastic discretization.

We conclude this introductory part by noting that the problem (1.2) is well-posed. This follows trivially from (1.1) and the well-posedness of the deterministic diffusion problem (see also, e.g. Todor, 2005).

¹Here and throughout the paper, linear complexity is understood as log-linear with respect to the number of degrees of freedom for a FE discretization of one deterministic version of the stochastic BVP.

THEOREM 1.3 If Assumption 1.1 holds and $p \geq 0$, then for any $f \in L^p(\Omega, H^{-1}(D))$, there exists a unique $u \in L^p(\Omega, H_0^1(D))$ that is a solution to (1.2) (here $p = 0$ corresponds to measurability). Moreover, for $p \geq 1$ there exists a positive constant c_a , independent of f and u , such that

$$\|u\|_{L^p(\Omega, H_0^1(D))} \leq c_a \|f\|_{L^p(\Omega, H^{-1}(D))}.$$

2. Separation of deterministic and stochastic variables

To reduce (1.2) to a high-dimensional deterministic problem, we separate the deterministic and stochastic variables in the coefficient $a(x, \omega)$ using an expansion in a deterministic basis, with random coefficients. Several choices are possible here, of which we mention and discuss the Legendre and KL expansions. We consider a splitting of the diffusion coefficient into a *deterministic expectation* e and a *random fluctuation* r . To this end, we must strengthen the positivity Assumption 1.1.

ASSUMPTION 2.1 The random field $a \in L^\infty(D \times \Omega)$ satisfying (1.1) can be represented as

$$a(x, \omega) = e(x) + r(x, \omega) \quad \forall (x, \omega) \in D \times \Omega, \quad (2.1)$$

with a positive $e \in L^\infty(D)$ (not necessarily equal to the mean field E_a) such that

$$0 < e_- \leq e(x) \leq e_+ < \infty \quad \forall x \in D. \quad (2.2)$$

It follows from (2.2) that $r \in L^\infty(D \times \Omega)$ too, and we require that the fluctuation r be pointwise smaller than the deterministic expectation e .

ASSUMPTION 2.2 For the representation (2.1), we assume that

$$0 \leq \sigma := \operatorname{ess\,sup}_{x \in D} \frac{\|r(x, \cdot)\|_{L^\infty(\Omega)}}{e(x)} < 1. \quad (2.3)$$

REMARK 2.3 The constant expectation

$$e(x) := (a_- + a_+)/2 \quad \forall x \in D,$$

satisfies Assumption 2.2 with $\sigma \leq (a_+ - a_-)/(a_+ + a_-) < 1$.

The more natural (from a statistical point of view) choice $e = E_a$ satisfies (2.3) if the probability density of $r(x, \cdot)$ is symmetric for any $x \in D$, i.e. if positive and negative fluctuations occur with equal probabilities.

Concerning the fluctuation term r , we also formulate a modelling assumption as well as a condition of regularity in the physical variable.

ASSUMPTION 2.4 The fluctuation r can be represented in $L^\infty(D \times \Omega)$ as a convergent series

$$r = \sum_{m=1}^{\infty} \psi_m \otimes X_m \quad (2.4)$$

with known deterministic $\psi_m \in L^\infty(D)$ and stochastic $X_m \in L^\infty(\Omega)$. Without loss of generality we also assume that $\psi_m, X_m \neq 0$ for all $m \in \mathbb{N}_+$.

The representation formula (2.4) describes the tensor-product nature of the random field r , and achieves the separation of the deterministic and stochastic variables, $x \in D$ and $\omega \in \Omega$, respectively. Note also that we require uniform convergence of (2.4) in order to allow control of the error in the

solution to (1.2) via Strang's lemma, after truncation of (2.4). The regularity of the random field r is quantified by the convergence rate of the series (2.4).

ASSUMPTION 2.5 The fluctuation r admits a representation (2.4) for which there exist constants $c_r, c_{1,r}, \kappa > 0$ such that

$$\|\psi_m \otimes X_m\|_{L^\infty(D \times \Omega)} \leq c_r \exp(-c_{1,r} m^\kappa) \quad \forall m \in \mathbb{N}_+. \quad (2.5)$$

In the following two sections, Assumption 2.5 will be shown to hold with $\kappa = 1/d$ if the fluctuation r is piecewise analytic in the physical variable $x \in D \subset \mathbb{R}^d$. Two examples of separating expansions (2.4) will be presented and discussed in detail, the Legendre and the KL expansions, respectively.

We further assume that complete probabilistic information on the stochastic part of the separating expansion (2.4) is available, as follows.

ASSUMPTION 2.6 The joint probability density functions of the family $\mathcal{X} := (X_m)_{m \in \mathbb{N}_+}$ are known.

In fact, this assumption will be only needed later for the postprocessing of the chaos solution to our stochastic problem (1.2).

2.1 Legendre expansion

The validity of Assumption 2.5 with $\kappa = 1/d$ and the existence of a Legendre expansion (2.4) follow from standard approximation theory of analytic functions (see, e.g. Davis, 1963), if the random fluctuation r is piecewise analytic in the physical variable, with values in $L^\infty(\Omega)$ ($r \in \mathcal{A}_{\text{pw}}(\overline{D}, L^\infty(\Omega))$).

EXAMPLE 2.7 If $D \subset [-1, 1]^d$ and $r \in \mathcal{A}([-1, 1]^d, L^\infty(\Omega))$, then a representation (2.4) exists with $(\psi_m)_{m \in \mathbb{N}_+}$ being the Legendre polynomials in $[-1, 1]^d$ (tensor-products of standard Legendre polynomials in $[-1, 1]$ scaled to have L^2 -norm equal to 1) and

$$\forall n \in \mathbb{N}_+: \quad X_m(\omega) := \int_{[-1, 1]^d} r(x, \omega) \psi_m(x) dx, \quad P\text{-a.e. } \omega \in \Omega.$$

Moreover, Assumption 2.5 holds with $\kappa := 1/d$ and $c_{1,r}$ depending on the size of the analyticity domain of r in a complex neighbourhood of $[-1, 1]^d$.

2.2 KL expansion

An alternative to the Legendre expansion is the KL series, which is known to be the $L^2(D \times \Omega)$ optimal representation satisfying the separation ansatz (2.4) (see also Schwab & Todor, 2006). For analytic fluctuations r , the convergence rate of the KL series is also exponential, i.e. qualitatively similar to that of the Legendre expansion. However, determining it requires an additional eigenpair computation for the compact integral operator \mathcal{C}_r with kernel C_r given by the two-point correlation of r ,

$$C_r: D \times D \rightarrow \mathbb{R}, \quad C_r(x, x') := \int_{\Omega} r(x, \omega) r(x', \omega) dP(\omega), \quad (x, x') \in D \times D.$$

We start by noting that $\mathcal{C}_r: L^2(D) \rightarrow L^2(D)$ given by

$$(\mathcal{C}_r u)(x) := \int_D C_r(x, x') u(x') dx', \quad \lambda\text{-a.e. } x \in D, \quad \forall u \in L^2(D), \quad (2.6)$$

is a symmetric, nonnegative definite and compact integral operator. It therefore possesses a countable sequence $(\lambda_m, \phi_m)_{m \in \mathbb{N}_+}$ of eigenpairs with

$$\mathbb{R} \ni \lambda_m \searrow 0, \quad \text{as } m \nearrow \infty,$$

where the KL eigenvalues are enumerated in decreasing order of magnitude counting multiplicity. We also have the following result (see Loève, 1977).

THEOREM 2.8 Under Assumption 2.1, there exists a sequence $\mathcal{X} := (X_m)_{m \in \mathbb{N}_+}$ of uncorrelated (and centred at 0 if $e = E_a$) random variables

$$\int_{\Omega} X_n(\omega) X_m(\omega) dP(\omega) = \delta_{nm} \quad \forall n, m \in \mathbb{N}_+, \quad (2.7)$$

such that the random field r can be expanded in $L^2(D \times \Omega)$ as

$$r(x, \omega) = a(x, \omega) - e(x) = \sum_{m \in \mathbb{N}_+} \sqrt{\lambda_m} \phi_m(x) X_m(\omega). \quad (2.8)$$

Note that the $L^2(D \times \Omega)$ convergence of the KL expansion is due to the trace-class condition

$$\sum_{m=1}^{\infty} \lambda_m = \text{Tr}(\mathcal{C}_r) = \int_D \int_{\Omega} r(x, \omega)^2 < \infty. \quad (2.9)$$

REMARK 2.9 The convergence rate of the KL series in $L^2(D \times \Omega)$ is equal to the convergence rate of the trace in (2.9).

Note that the $L^2(D \times \Omega)$ convergence of the KL expansion (2.8) is not strong enough to allow control of the error in the solution of (1.2) via Strang's lemma, after truncation of (2.8). However, analytic regularity of r in the physical variable plus uniform boundedness of the family $\mathcal{X} = (X_m)_{m \in \mathbb{N}_+} \subset L^\infty(\Omega)$ will be next shown to ensure the uniform convergence of the KL expansion (2.8).

ASSUMPTION 2.10 The family $\mathcal{X} = (X_m)_{m \in \mathbb{N}_+}$ of random variables is uniformly bounded in $L^\infty(\Omega)$, i.e. there exists $c_{\mathcal{X}} > 0$ such that

$$\|X_m\|_{L^\infty(\Omega, dP)} \leq c_{\mathcal{X}} \in \mathbb{R} \quad \forall m \in \mathbb{N}_+. \quad (2.10)$$

The eigenvalue and eigenfunction decay estimates derived in Propositions 2.13 and 2.16 of the following two sections immediately imply the desired strong convergence result.

PROPOSITION 2.11 If $D \subset [-1, 1]^d$, $r \in \mathcal{A}([-1, 1]^d, L^\infty(\Omega))$ with the associated KL expansion given by (2.8), and Assumption 2.10 holds, then Assumption 2.5 holds too, with $\kappa := 1/d$ and $\psi_m := \sqrt{\lambda_m} \phi_m$ for all $m \in \mathbb{N}_+$.

For the proof, we refer to Schwab & Todor (2006), Theorem 3.5.

2.2.1 Eigenvalue decay. Next we state decay rates for the KL eigenvalues in terms of regularity of the correlation kernel C_r . The results we present in this section are standard (see, e.g. König, 1986; Pietsch, 1987; Pinkus, 1985), following from the abstract theory of Weyl/approximation/entropy numbers via

approximation of K by discrete, finite-rank (separable with respect to (x, x')) kernels. Roughly speaking, the smoother the kernel the faster the eigenvalue decay, with finite Sobolev regularity implying algebraic rates of decay and analytic kernels giving rise to exponential decay.

Remarkably, all these results hold for piecewise regular kernels on product subdomains of D , in the sense of Definition 2.12 below. Note that general piecewise regularity allowing singularities on the diagonal set of $D \times D$ ensure in general only a slower eigenvalue decay (see, e.g. König, 1986; Ghanem & Spanos, 1997, for examples with known exact eigenelements). We focus on the case of piecewise analytic correlation kernels C_r , and refer the reader to Todor (2006) for a discussion of less regular kernels.

DEFINITION 2.12 If D is a bounded domain in \mathbb{R}^d , a measurable function $C_r: D \times D \rightarrow \mathbb{R}$ is said to be ‘*piecewise analytic on*’ $D \times D$ if there exists a finite family $\mathcal{D} = (D_j)_{j \in \mathcal{J}}$ of subdomains of D such that

- i. $D_j \cap D_{j'} = \emptyset \forall j, j' \in \mathcal{J}$ with $j \neq j'$;
- ii. $D \setminus \cup_{j \in \mathcal{J}} D_j$ is a null set in \mathbb{R}^d ;
- iii. $\overline{D} \subset \cup_{j \in \mathcal{J}} \overline{D_j}$;
- iv. $C_r|_{D_j \times D_{j'}}$ is analytic on $\overline{D_j} \times \overline{D_{j'}} \forall j, j' \in \mathcal{J}$.

We denote by $\mathcal{A}_{\mathcal{D}}(D^2)$ the space of piecewise analytic functions on $D \times D$ in the sense given above. Moreover, if there exists also a finite family $\mathcal{G} = (G_j)_{j \in \mathcal{J}}$ of open sets in \mathbb{R}^d such that

- v. $\overline{D_j} \subset G_j \forall j \in \mathcal{J}$;
- vi. $C_r|_{D_j \times D_{j'}}$ has an analytic continuation to $G_j \times G_{j'} \forall j, j' \in \mathcal{J}$,

then we say that K is piecewise analytic on a covering of $D \times D$ and we denote by $\mathcal{A}_{\mathcal{D}, \mathcal{G}}(D^2)$ the corresponding space.

Similarly, we introduce spaces of piecewise analytic functions defined on D , which we denote by $\mathcal{A}_{\mathcal{D}}(D)$, $\mathcal{A}_{\mathcal{D}, G}(D)$, etc.

The next result, established in Proposition 2.18 in Schwab & Todor (2006), shows that the eigenvalues in the KL expansion decay exponentially provided the correlation kernel is piecewise analytic.

PROPOSITION 2.13 If $C_r \in \mathcal{A}_{\mathcal{D}, G}(D^2)$ and $(\lambda_m)_{m \in \mathbb{N}_+}$ is the eigenvalue sequence of its associated integral operator (2.6), then there exist constants $c_1, c_2 > 0$ such that

$$0 \leq \lambda_m \leq c_1 \exp(-c_2 m^{1/d}) \quad \forall m \in \mathbb{N}_+. \quad (2.11)$$

EXAMPLE 2.14 One is often interested in Gaussian kernels of the form

$$C_r(x, x') := \sigma^2 \exp(-|x - x'|^2 / (\gamma^2 \mathcal{A}^2)) \quad \forall (x, x') \in D \times D, \quad (2.12)$$

where $\sigma, \gamma > 0$ are real parameters (standard deviation and correlation length, respectively) and \mathcal{A} is the diameter of the domain D . C_r given by (2.12) has an entire continuation to \mathbb{C}^d and defines a nonnegative compact operator via (2.6).

Since C_r given by (2.12) admits an analytic continuation to the whole complex space \mathbb{C}^d , the eigenvalue decay is in this case even faster than in (2.11).

PROPOSITION 2.15 If C_r is given by (2.12), then for the eigenvalue sequence $(\lambda_m)_{m \in \mathbb{N}_+}$ of the corresponding integral operator \mathcal{C}_r defined by (2.6) we have

$$0 \leq \lambda_m \leq c_{\sigma, \gamma} \frac{(1/\gamma)^{m^{1/d}}}{\Gamma(m^{1/d}/2)} \quad \forall m \in \mathbb{N}_+. \quad (2.13)$$

For a proof, we refer to Proposition 2.19 in Schwab & Todor (2006).

Note that the decay estimate (2.13) is subexponential in dimension $d > 1$, and this is essentially due to the higher multiplicity of the eigenvalues in dimension larger than 1 (this can be explicitly seen, e.g. for the separable kernel (2.12) on a product domain D).

2.2.2 Eigenfunction estimates. The smoothness assumption on the correlation kernel C_r allows also a good control of the eigenfunctions in terms of corresponding eigenvalues via the Gagliardo–Nirenberg inequalities. For a proof of the following result, we refer the reader to Todor (2006).

PROPOSITION 2.16 Let $C_r \in L^2(D \times D)$ be piecewise analytic on $D \times D$, such that all subdomains D_j in Definition 2.12 satisfy the uniform cone property. Denote by $(\lambda_m, \phi_m)_{m \in \mathbb{N}_+}$ the sequence of eigenpairs of the associated integral operator via (2.6), such that $\|\phi_m\|_{L^2(D)} = 1$ for all $m \in \mathbb{N}_+$. Then for any $s > 0$ and any multi-index $\alpha \in \mathbb{N}^d$, there exists $c_{r, \alpha, s} > 0$ such that

$$\|\partial^\alpha \phi_m\|_{L^\infty(D)} \leq c_{r, \alpha, s} |\lambda_m|^{-s} \quad \forall m \in \mathbb{N}_+. \quad (2.14)$$

3. Uncertainty parameterization

Throughout this section, we suppose that the separating expansion (2.4) of the random fluctuation r satisfies the decay Assumption 2.5. As shown before, this is the case if r is piecewise analytic in the physical variable $x \in D$ (in which case $\kappa = 1/d$).

3.1 Truncation of fluctuation expansion

Since computations can handle only finite data sets, we truncate the fluctuation expansion (2.4) and introduce, for any $M \in \mathbb{N}$, the truncated stochastic diffusion coefficient

$$a_M(x, \omega) = e(x) + \sum_{m=1}^M \psi_m(x) X_m(\omega), \quad (3.1)$$

for which the following pointwise error estimate holds due to Assumption 2.5.

PROPOSITION 3.1 If Assumption 2.5 holds, then

$$\|a - a_M\|_{L^\infty(D \times \Omega)} \leq c_r \exp(-c_{1,r} M^\kappa) \quad \forall M \in \mathbb{N}. \quad (3.2)$$

The diffusion problem with truncated coefficient a_M is therefore well-posed for M large enough (depending on a). This follows immediately from Strang’s lemma, which allows also explicit control of the error in the solution u to (1.2).

COROLLARY 3.2 If the stochastic diffusion coefficient a satisfies Assumptions 2.1 and 2.5, then there exists a truncation order $M_{a,r} \in \mathbb{N}$ of the expansion (2.4) such that (3.4) below is well-posed in

$L^\infty(\Omega, H_0^1(D))$ for any $M \geq M_{a,r}$. Moreover, if u and u_M are the unique solutions in $L^\infty(\Omega, H_0^1(D))$ of

$$-\operatorname{div}(a(\cdot, \omega) \nabla u(\cdot, \omega)) = f(\cdot) \quad \text{in } H^{-1}(D), \quad P\text{-a.e. in } \Omega, \quad (3.3)$$

and

$$-\operatorname{div}(a_M(\cdot, \omega) \nabla u_M(\cdot, \omega)) = f(\cdot) \quad \text{in } H^{-1}(D), \quad P\text{-a.e. in } \Omega, \quad (3.4)$$

respectively, then

$$\|u - u_M\|_{L^\infty(\Omega, H_0^1(D))} \leq c_{a,r} \exp(-c_{1,r} M^\kappa) \cdot \|u\|_{L^\infty(\Omega, H_0^1(D))} \quad (3.5)$$

for all $M \geq M_{a,r}$.

For a proof, we refer to Todor (2005).

REMARK 3.3 If the expectation e is chosen to be equal to the mean-field E_a and the family $\mathcal{X} = (X_m)_{m \in \mathbb{N}_+}$ is assumed to be independent, then (3.4) is well-posed for any $M \geq 0$, i.e. $M_{a,r}$ can be chosen equal to 0 in Corollary 3.2. The possible loss of ellipticity in (3.4)—due to a Gibbs' effect—is therefore not possible in the presence of an independent family $\mathcal{X} = (X_m)_{m \in \mathbb{N}_+}$, even in the case of slow, nonuniform convergence of the separating expansion (2.4). The typical example here is the KL expansion of a fluctuation r with low regularity of its two-point correlation C_r , which exhibits only slow convergence in $L^2(D \times \Omega)$.

Under the assumptions in Remark 3.3, the well-posedness of (3.4) can be seen for instance for the KL expansion as follows. For any $N \in \mathbb{N}_+$ denote by $\Sigma_N \subset \Sigma$ the σ -algebra generated by the random variables X_1, X_2, \dots, X_N . For any $M > N$, from (3.1) it follows (conditional expectations) that

$$\mathbb{E}[a_M | \Sigma_N] = a_N, \quad (3.6)$$

since $(X_m)_{m \in \mathbb{N}_+}$ are assumed to be independent and, by construction of the KL expansion, centred at 0.

For any $\Omega_N \in \Sigma_N$, we use (3.6) and the defining property of the conditional expectation to write

$$\begin{aligned} \int_D \left(\int_{\Omega_N} (a_N - a) dP(\omega) \right)^2 dx &= \int_D \left(\int_{\Omega_N} (a_M - a) dP(\omega) \right)^2 dx \\ &\leq \int_D \int_{\Omega_N} (a_M - a)^2 dP(\omega) dx \xrightarrow{M \rightarrow \infty} 0, \end{aligned} \quad (3.7)$$

due to $a_M \rightarrow a$ in $L^2(D \times \Omega)$ as $M \nearrow \infty$. Since $\Omega_N \in \Sigma_N$ was arbitrary, we conclude from (3.7) that

$$a_N = \mathbb{E}[a | \Sigma_N].$$

The positivity of the conditional expectation ensures then that the lower and upper bounds on a hold also for a_N .

3.2 Parametric deterministic problem

In this section, we connect (3.4) obtained by truncation at level $M \in \mathbb{N}$ of the separating expansion (2.4) of the random fluctuation r to an auxiliary, purely deterministic parametric problem. Without loss of generality, we suppose in the following that for $(X_m)_{m \in \mathbb{N}_+}$ in (2.4) we have (this can be achieved by a

rescaling of ψ_m and X_m)

$$\|X_m\|_{L^\infty(\Omega)} = 1/2 \quad \forall m \in \mathbb{N}_+, \quad (3.8)$$

so that

$$\text{Ran } X_m \subseteq I := [-1/2, 1/2] \quad \forall m \in \mathbb{N}_+.$$

To a_M we associate the function $\tilde{a}_M : D \times I^M \rightarrow \mathbb{R}$ defined by

$$\tilde{a}_M(x, y_1, y_2, \dots, y_M) := e(x) + \sum_{m=1}^M \psi_m(x) y_m \quad (3.9)$$

for all $y = (y_1, y_2, \dots, y_M) \in I^M$ and $x \in D$.

We now consider the purely deterministic, parametric elliptic problem of finding $\tilde{u}_M : I^M \rightarrow H_0^1(D)$ such that

$$-\text{div}(\tilde{a}_M(\cdot, y) \nabla \tilde{u}_M(\cdot, y)) = f(\cdot) \quad \text{in } H^{-1}(D) \quad \forall y \in I^M. \quad (3.10)$$

The uniform ellipticity of all truncates a_M for $M \geq M_{a,r}$, following from Corollary 3.2, ensures the well-posedness of (3.10). The solution of (3.4) can be obtained from the solution of (3.10) by back substitution, as follows.

PROPOSITION 3.4 If \tilde{u}_M is the solution of (3.10) and u_M solves (3.4), then

$$u_M(x, \omega) = \tilde{u}_M(x, X_1(\omega), X_2(\omega), \dots, X_M(\omega)), \quad (3.11)$$

$(\lambda \times P)$ -a.e. $(x, \omega) \in D \times \Omega$.

The proof is immediate, observing that both the left-hand side and the right-hand side of (3.11) solve the well-posed problem (3.4).

Assuming that sufficient statistical information is available on the family $\mathcal{X} = (X_m)_{m \in \mathbb{N}_+}$ to allow the postprocessing (i.e. the computation of various statistics of u_M , see Assumption 2.6) via (3.11), Proposition 3.4 reduces the elliptic problem with stochastic data (1.2) to a question in *approximation theory* for the parametric (in $y \in I^M$) solution to (3.10), which we formulate as follows.

PROBLEM 3.5 For any M , compute the solution \tilde{u}_M to (3.10) in $L^\infty(I^M, H_0^1(D))$ up to an error of $\exp(-c_{2,r} M^\kappa)$.

Note that the truncation order M of the separating expansion (2.4) is the dimension of the parameter space I^M and, in fact, a discretization parameter. In Section 4, the aim will be therefore to solve Problem 3.5 by developing efficient approximations for \tilde{u}_M as a function of $y \in I^M$. The key point of our analysis will be the regularity of \tilde{u}_M with respect to the *stochastic parameter* y , to which we shall refer as *stochastic regularity*. While it is easy to see that the dependence on y of the solution \tilde{u}_M is analytic, we shall prove that the domain of analyticity of \tilde{u}_M as a function of y_m increases in size as $m \nearrow \infty$. Our estimates indicate in particular that \tilde{u}_M as a function of $y_m \in I$ becomes ‘flat’ as m increases at a rate which is governed by the convergence rate of the separating expansion (2.4).

To see this, we note that the decay rate of the expansion (2.4) of the random fluctuation r shows the decreasing sensitivity of \tilde{a}_M with respect to y_m as $m \nearrow M$. Intuitively, \tilde{u}_M is then expected to exhibit a similar behaviour.

Note that we are not interested in approximating \tilde{u}_M with arbitrarily high accuracy, but only up to an error which matches the truncation error $O(\exp(-c_{1,r}M^k))$ in the separating expansion (2.4). The required accuracy thus depends on the dimension M of the domain I^M over which the function \tilde{u}_M to be approximated is defined.

4. Sparse Galerkin approximation

For the solution of the approximation problem 3.5, we use an approach inspired by nonlinear approximation results. To describe it, let $(\phi_{M,\alpha})_{\alpha \in A} \subset L^\infty(I^M)$ be a family of real-valued functions defined on the hypercube I^M such that \tilde{u}_M admits the expansion

$$\tilde{u}_M = \sum_{\alpha \in A} \phi_{M,\alpha} \otimes c_{M,\alpha} \quad \text{in } L^\infty(I^M, H_0^1(D)), \quad (4.1)$$

with $c_{M,\alpha} \in H_0^1(D)$ for all $\alpha \in A$.

DEFINITION 4.1 If (4.1) holds, we call the series on the right-hand side of (4.1) a chaos expansion of dimension M of u , the solution to (1.2).

For a finite index set $A' \subset A$, we define the corresponding truncation of (4.1)

$$\tilde{u}_{M,A'} := \sum_{\alpha \in A'} \phi_{M,\alpha} \otimes c_{M,\alpha} \in L^\infty(I^M, H_0^1(D)). \quad (4.2)$$

In the spirit of the theory of adaptive/best N -term approximation, we consider the most economical chaos truncation (4.2) which achieves an accuracy comparable with that obtained after truncation of the separating expansion of r (see Corollary 3.2).

DEFINITION 4.2 If (4.1) holds, we define

$$A_M := \operatorname{argmin} \left\{ |A'| : \|\tilde{u}_M - \tilde{u}_{M,A'}\|_{L^\infty(I^M, H_0^1(D))} \leq \|u - u_M\|_{L^\infty(\Omega, H_0^1(D))} \right\} \quad (4.3)$$

and call the truncation \tilde{u}_{M,A_M} the adapted chaos expansion of dimension M of u , the solution to (1.2).

Due to explicit control of the truncation error in the case of a diffusion coefficient with known decay rate of coefficients in the fluctuation expansion (2.4), it is more convenient to work with the following more explicit version of Definition 4.2.

DEFINITION 4.3 If Assumption 2.5 and (4.1) hold, we set

$$A_M := \operatorname{argmin} \left\{ |A'| : \|\tilde{u}_M - \tilde{u}_{M,A'}\|_{L^\infty(I^M, H_0^1(D))} \leq \exp(-c_{1,r}M^k) \right\} \quad (4.4)$$

and call the truncation \tilde{u}_{M,A_M} the adapted chaos expansion of dimension M of u , the solution to (1.2).

The aim of the following sections is the (approximate) identification of the index set Σ_M , based on the regularity properties of \tilde{u}_M with respect to y (analyticity and explicit bounds for all derivatives), if the family $(\phi_{M,\alpha})_{\alpha \in A} \subset L^\infty(I^M)$ is chosen to generate the standard FE spaces (piecewise polynomials of fixed degree on regular meshes/polynomials), corresponding to h - and p -FE over I^M , respectively.

4.1 Stochastic regularity

We start by observing that Assumption 2.5 and (3.8) trivially ensure the following norm estimates

$$0 \leq \rho_m := \|\psi_m\|_{L^\infty(D)} \leq c_r \exp(-c_{1,r} m^\kappa) \quad \forall m \in \mathbb{N}_+. \quad (4.5)$$

Explicit bounds on all derivatives of \tilde{u}_M are then obtained by either using (3.10), (3.9) (as shown below) or by Cauchy's formula (see, e.g. Todor, 2005).

PROPOSITION 4.4 If \tilde{u}_M solves (3.10), then

$$\|\partial_y^\alpha \tilde{u}_M(y, \cdot)\|_{H_0^1(D)} \leq c_{a,r}^{|\alpha|} \cdot |\alpha|! \cdot \prod_{m=1}^M \rho_m^{\alpha_m} \cdot \|\tilde{u}_M(y, \cdot)\|_{H_0^1(D)} \quad (4.6)$$

$\forall y \in I^M, \forall \alpha \in \mathbb{N}^M, \forall M \in \mathbb{N}, M \geq M_{a,r}$.

Proof. We prove the estimate (4.6) by induction on $|\alpha|$. Since (4.6) is clear for $|\alpha| = 0$, we assume it to hold also for all $\alpha \in \mathbb{N}^M$ such that $|\alpha| \leq k$, for some $k \in \mathbb{N}$. We consider a multi-index α such that $|\alpha| = k + 1$ and we apply ∂_y^α to (3.10). We obtain

$$-\operatorname{div}(\tilde{a}_M(\cdot, y) \nabla \partial_y^\alpha \tilde{u}_M(\cdot, y)) = \sum_{m=1}^M \alpha_m \operatorname{div}(\psi_m(\cdot) \nabla \partial_y^{\alpha - e_m} \tilde{u}_M(\cdot, y))$$

from which it follows that

$$c_{a,r} \|\partial_y^\alpha \tilde{u}_M(\cdot, y)\|_{H_0^1(D)} \leq \sum_{m=1}^M \alpha_m \rho_m \|\partial_y^{\alpha - e_m} \tilde{u}_M(\cdot, y)\|_{H_0^1(D)}. \quad (4.7)$$

The desired estimate then follows by using (4.6) in (4.7) for all multi-indices $\alpha - e_m$, $1 \leq m \leq M$, whose length equals k . \square

4.2 Sparse chaos approximations

Using Proposition 4.4, we next investigate convergence rates of adapted chaos approximations for $\tilde{u}_M: I^M \rightarrow H_0^1(D)$, if tensor-product families $(\phi_{M,\alpha})_{\alpha \in \mathcal{A}} \subset L^\infty(I^M, H_0^1(D))$ corresponding to standard h - or p -FE in I^M are chosen in (4.1) to expand \tilde{u}_M .

4.2.1 h -FE-based adapted approximation. For $p \in \mathbb{N}_+$ and $l \in \mathbb{N}$, let $V^{l,p}$ be the space of piecewise polynomials of degree at most $p - 1$ on a regular mesh of size 2^{-l} in I . We set $V^{-1,p} := \{0\}$, and by

$$W^{l,p} := V^{l,p} \cap (V^{l-1,p})^\perp,$$

we define the hierarchical excess of the scale $(V^{l,p})_{l \in \mathbb{N}}$, where the orthogonal complement is taken in the sense of $L^2(I)$. In this way, we obtain an $L^2(I)$ -orthogonal decomposition

$$L^2(I) = \bigoplus_{l=0}^{\infty} W^{l,p}. \quad (4.8)$$

If H is an arbitrary Hilbert space and P_V denotes the $L^2(I, H) \simeq L^2(I) \otimes H$ projection onto the closed subspace $V \otimes H$ of $L^2(I)$, the standard (vector-valued) approximation property of the scale $(V^{l,p})_{l \in \mathbb{N}}$ reads

$$\|v - P_{V^{l,p}} v\|_{L^2(I, H)} \leq c_p 2^{-lp} \|\partial^p v\|_{L^2(I, H)} \quad \forall v \in H^p(I, H), \quad (4.9)$$

with some constant $c_p > 0$.

REMARK 4.5 Note that an estimate similar to (4.9) holds also in the $L^\infty(I, H)$ norm, for $v \in W^{p,\infty}(I, H)$.

Using the FE spaces $V^{l,p}$ in I , we build FE spaces in I^M as tensor products. More precisely, for any multi-index $\mathbf{l} = (l_1, l_2, \dots, l_M) \in \mathbb{N}^M$, we define

$$W^{\mathbf{l},p} := \bigotimes_{m=1}^M W^{l_m,p},$$

which enables us via (4.8) to decompose $L^2(I^M)$ as

$$L^2(I^M) = \bigoplus_{\mathbf{l} \in \mathbb{N}^M} W^{\mathbf{l},p}. \quad (4.10)$$

In $L^2(I^M, H)$, we then have

$$v = \sum_{\mathbf{l} \in \mathbb{N}^M} v^{\mathbf{l}}, \quad v^{\mathbf{l}} := P_{W^{\mathbf{l},p} \otimes H} v \quad \forall v \in L^2(I^M, H). \quad (4.11)$$

For $\tilde{u}_M \in L^\infty(I^M, H_0^1(D)) \subset L^2(I^M, H_0^1(D))$, the solution to (3.10), we next estimate the size of the general term (detail of \tilde{u}_M at level \mathbf{l}) in the corresponding orthogonal decomposition (4.11) with $H := H_0^1(D)$. To this end, we first introduce several pieces of notation. We define the length $|\mathbf{l}|$ of a multi-index $\mathbf{l} = (l_1, l_2, \dots, l_M) \in \mathbb{N}^M$ by

$$|\mathbf{l}| := l_1 + l_2 + \dots + l_M. \quad (4.12)$$

Further, the support of \mathbf{l} will be denoted by

$$\mathcal{J}_1 := \text{supp}(\mathbf{l}) = \{m: 1 \leq m \leq M, l_m > 0\}, \quad (4.13)$$

and its length by $j_1 := |\mathcal{J}_1|$, so that $\mathcal{J}_1 = \{m_1, m_2, \dots, m_{j_1}\}$.

PROPOSITION 4.6 If \tilde{u}_M solves (3.10) and Assumption 2.5 holds, then

$$\|\tilde{u}_M^{\mathbf{l}}\|_{L^2(I^M, H_0^1(D))} \leq c_{a,p}^{j_1} \cdot 2^{-|\mathbf{l}|p} \cdot (p j_1)! \cdot \prod_{j=1}^{j_1} \rho_{m_j}^p \cdot \|\tilde{u}_M\|_{L^2(I^M, H_0^1(D))}, \quad (4.14)$$

where $\tilde{u}_M^{\mathbf{l}} := P_{W^{\mathbf{l},p} \otimes H_0^1(D)} \tilde{u}_M$ for all $\mathbf{l} \in \mathbb{N}^M$.

Proof. For a fixed multi-index $\mathbf{l} \in \mathbb{N}^M$, we define its support multi-index $\mathbf{e} := (e_1, e_2, \dots, e_M) \in \mathbb{N}^M$ (depending on \mathbf{l}) by

$$e_m := \begin{cases} 1 & \text{if } l_m > 0, \\ 0 & \text{if } l_m = 0, \end{cases} \quad \forall 1 \leq m \leq M,$$

and write

$$\tilde{u}_M^{\mathbf{l}} = P_{W^{\mathbf{l},p}} \tilde{u}_M = \bigotimes_{m=1}^M (P_{V^{l_m,p}} - P_{V^{l_m-1,p}}) \tilde{u}_M.$$

Replacing $P_{V^{l_m,p}} - P_{V^{l_m-1,p}}$ by $P_{V^{l_m,p}} - I + I - P_{V^{l_m-1,p}}$ for all m in the support of \mathbf{l} and expanding the resulting product, we obtain

$$\tilde{u}_M^{\mathbf{l}} = \sum_{\mathbf{f} \in \mathbb{N}^M, \mathbf{f} \leq \mathbf{e}} (-1)^{M-|\mathbf{f}|} \left(\bigotimes_{m=1}^M Q_{l_m, f_m} \right) \tilde{u}_M, \quad (4.15)$$

where

$$Q_{l_m, f_m} := \begin{cases} P_{V^{0,p}} & \text{if } l_m = 0, \\ I - P_{V^{l_m-f_m,p}} & \text{if } l_m > 0. \end{cases}$$

Using the approximation property (4.9) and noting that the sum in (4.15) consists of $2^{\dot{\mathbf{l}}}$ terms, we deduce that

$$\begin{aligned} \|\tilde{u}_M^{\mathbf{l}}\|_{L^2(I^M, H_0^1(D))} &\leq \sum_{\mathbf{f} \in \mathbb{N}^M, \mathbf{f} \leq \mathbf{e}} c_p^{\dot{\mathbf{l}}} 2^{-(|\mathbf{l}|-|\mathbf{f}|)p} \cdot \|\partial_y^{p \cdot \mathbf{e}} \tilde{u}_M\|_{L^2(I^M, H_0^1(D))} \\ &\leq \sum_{\mathbf{f} \in \mathbb{N}^M, \mathbf{f} \leq \mathbf{e}} (2^p c_p)^{\dot{\mathbf{l}}} 2^{-|\mathbf{l}|p} \cdot \|\partial_y^{p \cdot \mathbf{e}} \tilde{u}_M\|_{L^2(I^M, H_0^1(D))} \\ &\leq (2^{p+1} c_p)^{\dot{\mathbf{l}}} 2^{-|\mathbf{l}|p} \cdot \|\partial_y^{p \cdot \mathbf{e}} \tilde{u}_M\|_{L^2(I^M, H_0^1(D))}. \end{aligned} \quad (4.16)$$

Proposition 4.4 coupled with (4.16) leads now to the desired estimate (4.14). \square

REMARK 4.7 Based on Remark 4.5, an estimate similar to (4.14) can be shown also in the $L^\infty(I^M, H_0^1(D))$ norm.

Next, we define a scale of sparse FE spaces in I^M which will be shown to achieve an almost optimal convergence rate of the corresponding approximations of \tilde{u}_M . In fact, we prescribe an index set in \mathbb{N}^M corresponding intuitively to the largest details in the orthogonal decomposition (4.11) of \tilde{u}_M . To this end, we introduce for $\mu, \nu \in \mathbb{N}$ the set of all multi-indices $\mathbf{l} \in \mathbb{N}^M$ not exceeding μ in length and having at most ν nontrivial entries,

$$A_{\mu, \nu} \subset \mathbb{N}^M, \quad A_{\mu, \nu} := \{\mathbf{l} \in \mathbb{N}^M : |\mathbf{l}| \leq \mu, \dot{\mathbf{l}} \leq \nu\}. \quad (4.17)$$

Correspondingly, we define, in view of (4.10), the following finite-dimensional subspace of $L^2(I^M)$,

$$\hat{V}^{\mu, \nu} := \bigoplus_{\mathbf{l} \in \Sigma_{\mu, \nu}} W^{\mathbf{l}, p}.$$

Using $\hat{V}^{\mu, \nu} \otimes H_0^1(D) \subset L^2(I, H_0^1(D))$ as semidiscretization space to approximate \tilde{u}_M , we now prove the main approximation result of this section. Here and in the following, $P_{\hat{V}^{\mu, \nu}}$ denotes the $L^2(I, H_0^1(D))$ projection onto $\hat{V}^{\mu, \nu} \otimes H_0^1(D)$.

PROPOSITION 4.8 If $\mu, \nu \in \mathbb{N}$ and Assumption 2.5 holds, then for \tilde{u}_M , the solution to (3.10), we have

$$\begin{aligned} \|\tilde{u}_M - P_{\hat{V}_{\mu,\nu}} \tilde{u}_M\|_{L^2(I^M, H_0^1(D))} &\leq c_{a,r,p,\theta} \left(e^{-\nu^{1+\kappa} c_{1,r} p / 2(1+\kappa)} \right. \\ &\quad \left. + 2^{-p(\mu+1)} \cdot e^{\nu \log(M+1) + \nu \log(\mu+2)} \right) \cdot \|\tilde{u}_M\|_{L^2(I^M, H_0^1(D))}. \end{aligned} \quad (4.18)$$

Besides,

$$\dim \hat{V}^{\mu,\nu} \leq p(M+1)^\nu (\mu+1)^{\nu+1} 2^\mu. \quad (4.19)$$

Proof. For notational ease and since in the following arguments all functions are evaluated in the standard norm of $L^2(I^M, H_0^1(D))$, we drop the corresponding subscript from all estimates. For arbitrary $\mu, \nu \in \mathbb{N}$, we write

$$\|\tilde{u}_M - P_{\hat{V}_{\mu,\nu}} \tilde{u}_M\| \leq \sum_{\mathbf{l} \in \mathbb{N}^M \setminus \Sigma_{\mu,\nu}} \|\tilde{u}_M^{\mathbf{l}}\| = \sum_{\substack{\mathbf{l} \in \mathbb{N}^M \\ |\mathbf{l}| > \nu}} \|\tilde{u}_M^{\mathbf{l}}\| + \sum_{\substack{\mathbf{l} \in \mathbb{N}^M \\ |\mathbf{l}| \leq \nu \\ |\mathbf{l}| > \mu}} \|\tilde{u}_M^{\mathbf{l}}\| \quad (4.20)$$

and next estimate the two sums S_1, S_2 on the right-hand side of (4.20) separately. In both cases, we use Proposition 4.6 and the notations (4.12), (4.13). We start with S_1 and write

$$S_1 = \sum_{j=\nu+1}^M \sum_{\substack{\mathbf{l} \in \mathbb{N}^M \\ |\mathbf{l}|=j}} \|\tilde{u}_M^{\mathbf{l}}\| \stackrel{(4.14)}{\leq} \sum_{j=\nu+1}^M c_{a,p}^j \cdot (pj)! \cdot \sum_{\substack{\mathbf{l} \in \mathbb{N}^M \\ |\mathbf{l}|=j}} 2^{-|\mathbf{l}|p} \cdot \prod_{k=1}^j \rho_{m_k}^p \cdot \|\tilde{u}_M\|. \quad (4.21)$$

Indexing the multi-indices in the second sum on the right-hand side of (4.21) over their support, we have that

$$\begin{aligned} \sum_{\substack{\mathbf{l} \in \mathbb{N}^M \\ |\mathbf{l}|=j}} 2^{-|\mathbf{l}|p} \cdot \prod_{k=1}^j \rho_{m_k}^p &\stackrel{(4.5)}{\leq} c_r^j \cdot \sum_{1 \leq m_1 < \dots < m_j \leq M} \prod_{k=1}^j e^{-c_{1,r} m_k^\kappa p} \cdot \sum_{l_{m_1}, \dots, l_{m_j}=1}^{\infty} 2^{-p(l_{m_1} + \dots + l_{m_j})} \\ &\leq \sum_{j=\nu+1}^M c_r^j \cdot \sum_{1 \leq m_1 < \dots < m_j \leq M} \prod_{k=1}^j e^{-c_{1,r} m_k^\kappa p}. \end{aligned} \quad (4.22)$$

We then use Lemma A.2 (with $y = c_{1,r} p$ and $z = (1 + \kappa)\theta p$) in (4.22) to obtain from (4.21),

$$S_1 \leq c_{r,p,\theta} \sum_{j=\nu+1}^M c_{a,r,p}^j \cdot (pj)! \cdot e^{-j^{1+\kappa}\theta p} \cdot \|\tilde{u}_M\|, \quad (4.23)$$

for any $\theta \in (0, c_{1,r}/(1 + \kappa))$. The fast, supergeometrical decay of the third factor on the right-hand side of (4.23) as $j \nearrow \infty$ (due to $\kappa > 0$) allows us to absorb of the first two (exponential and factorial). We conclude

$$S_1 \leq c_{a,r,p,\theta} e^{-\nu^{1+\kappa}\theta p} \cdot \|\tilde{u}_M\| \quad \forall \theta \in (0, c_{1,r}/(1 + \kappa)). \quad (4.24)$$

We now turn to the second sum S_2 in (4.20). Using again Proposition 4.6 and Lemma A.2, we similarly deduce that

$$\begin{aligned} S_2 &\stackrel{(4.14)}{\leq} \sum_{\substack{\mathbf{l} \in \mathbb{N}^M \\ j \leq v \\ |\mathbf{l}| > \mu}} c_{a,p}^j \cdot 2^{-|\mathbf{l}|p} \cdot (pj)! \cdot \prod_{j=1}^j \rho_{m_j}^p \cdot \|\tilde{u}_M\| \\ &\leq c_{a,r,p,\theta} \sum_{\substack{\mathbf{l} \in \mathbb{N}^M \\ j \leq v \\ |\mathbf{l}| > \mu}} e^{-j^{1+\kappa}\theta p} \cdot 2^{-|\mathbf{l}|p} \cdot \|\tilde{u}_M\|, \end{aligned} \quad (4.25)$$

for any $\theta \in (0, c_{1,r}/(1+\kappa))$. Now, using a counting argument on the right-hand side of (4.25) and then Lemma A.1 with $t = 2^{-p}$, we obtain

$$\begin{aligned} S_2 &\leq c_{a,r,p,\theta} \sum_{j=1}^v \binom{M}{j} e^{-j^{1+\kappa}\theta p} \cdot \sum_{l=\mu+1}^{\infty} \binom{l}{j} 2^{-pl} \cdot \|\tilde{u}_M\| \\ &\leq c_{a,r,p,\theta} 2^{-p(\mu+1)} \cdot \sum_{j=1}^v \binom{M}{j} e^{-j^{1+\kappa}\theta p} \cdot (1 - 2^{-p})^{-j-1} \cdot (\mu+2)^j \cdot \|\tilde{u}_M\| \\ &\leq c_{a,r,p,\theta} 2^{-p(\mu+1)} \cdot (M+1)^v \cdot (\mu+2)^v \cdot \|\tilde{u}_M\|, \end{aligned} \quad (4.26)$$

since $\binom{M}{j} \leq (M+1)^j$. The inequality (4.18) now follows from (4.24) and (4.26) by choosing $\theta = c_{1,r}/2(1+\kappa)$.

It remains to estimate the dimension of $\hat{V}^{\mu,v}$. Taking into account that the dimension of the detail space $W^{l,p}$ equals $p2^l$, we have

$$\begin{aligned} \dim \hat{V}^{\mu,v} &= p \sum_{q=0}^v \sum_{l=0}^{\mu} \binom{M}{q} \binom{l}{q} 2^l \leq p(M+1)^v \sum_{q=0}^v \sum_{l=0}^{\mu} \binom{l}{q} 2^l \\ &\leq p(M+1)^v \sum_{l=0}^{\mu} (l+1)^v 2^l \leq p(M+1)^v (\mu+1)^{v+1} 2^{\mu}, \end{aligned}$$

which concludes the proof. \square

COROLLARY 4.9 Under Assumption 2.5, there exist positive constants c_1, c_2 such that by choosing

$$\mu := \lceil c_1 M^{\kappa} \rceil, \quad v := \lceil c_2 M^{\kappa/(\kappa+1)} \rceil, \quad (4.27)$$

for \tilde{u}_M , the solution to (3.10), we have

$$\|\tilde{u}_M - P_{\hat{V}^{\mu,v}} \tilde{u}_M\|_{L^2(I^M, H_0^1(D))} \leq c_{a,r,p} \exp(-c_{1,r} M^{\kappa} + o(M^{\kappa})) \quad (4.28)$$

with κ as in Assumption 2.5 and with the number N of deterministic problems to be solved given by

$$N = \dim \hat{V}^{\mu,v} \leq c_{\kappa,p} \exp\left(\frac{c_{1,r}}{p} M^{\kappa} + o(M^{\kappa})\right), \quad (4.29)$$

as $M \nearrow \infty$, and with the same constant $c_{1,r}$ as in (3.5).

Proof. We choose the parameters in (4.18) of Proposition 4.8

$$\mu = \lceil (2(1+\kappa)/p)^{1/(1+\kappa)} M^{\kappa/(\kappa+1)} \rceil, \quad \nu = \lceil c_{1,r} M^\kappa / p \log 2 \rceil, \quad (4.30)$$

so that (4.28), (4.29) follow directly from (4.18) and (4.19), respectively. \square

REMARK 4.10 The proof of Corollary 4.9 also provides (see (4.30)) explicit values for the constants c_1, c_2 in (4.27). Note that c_1 depends only on κ, p and never exceeds 3, whereas c_2 scales linearly with $c_{1,r}$.

Combining (4.28) and (4.29), we reformulate the main approximation result of this section (optimality of the adapted chaos expansion) as follows.

THEOREM 4.11 If Assumption 2.5 holds, then

$$\inf_{v \in \hat{V}^{\mu,\nu} \otimes H_0^1(D)} \|\tilde{u}_M - v\|_{L^2(I^M, H_0^1(D))} \leq c_{a,r,p} N^{-p+o(1)} \quad \text{as } M \nearrow \infty, \quad (4.31)$$

and for the choice (4.27) of the parameters μ and ν , where $N = \dim \hat{V}^{\mu,\nu}$ is the number of deterministic diffusion problems in D to be solved.

REMARK 4.12 The convergence rate (4.31) of the h -FE-based adapted chaos expansion is, already for $p = 1$ (corresponding to piecewise constant elements), faster than the Monte Carlo ($O(N^{-1/2})$) or quasi-Monte Carlo ($O(N^{-1}(\log N)^{c_M})$) rate, where N denotes here the number of samples.

4.2.2 p -FE-based adapted approximation. The analyticity of \tilde{u}_M in $y \in I^M$, following, e.g. from Proposition 4.4, suggests the use of polynomial approximation (sometimes called *polynomial chaos*, corresponding to a polynomial basis $(\phi_{M,\alpha})_{\alpha \in A}$ in (4.1), see, e.g. Wiener, 1938) in the stochastic variable y . In this section we give, for any $M \in \mathbb{N}_+$, the construction of a polynomial space of low dimension in y , and in which $\tilde{u}^M: I^M \rightarrow H_0^1(D)$ can be approximated with the desired accuracy, i.e. up to an error of $O(e^{-c_{1,r} M^\kappa})$. The construction is based, just as in the case of h -FE discussed in Section 4.2.1, on a *priori* estimation of the coefficients $c_{M,\alpha}$ in (4.1) using a tensor-product basis in I^M . Selection of the largest estimated coefficients then leads to an upper estimate of the optimal index set Σ_M in Definition 4.3.

The tensor-product basis we use to represent \tilde{u}_M is given by the monomials in y_1, \dots, y_M ,

$$A := \mathbb{N}^M, \quad \phi_{M,\alpha}(y_1, y_2, \dots, y_M) := y_1^{\alpha_1} y_2^{\alpha_2} \cdots y_M^{\alpha_M} \quad \forall \alpha \in A.$$

The chaos expansion (4.1) then holds as the Taylor expansion of \tilde{u}_M around $y = 0$, due to the analyticity of the solution established in Proposition 4.4. Moreover, it can be shown (see also Todor, 2005) that \tilde{u}_M as a function of y admits a complex analytic extension to a cylindric complex neighbourhood $U^M \times i\mathbb{R}$ of I^M , where $I \subset U \subset \mathbb{R}$.

In analogy with the construction of the FE space $\hat{V}^{\mu,\nu}$ in Section 4.2.1, we consider, for $M, M', \eta, \mu, \nu \in \mathbb{N}$ with $M' \leq M$ and in the context of the p -FE, the polynomial space $\mathcal{P}_{M',\eta,\mu,\nu}$ in the M variables y_1, y_2, \dots, y_M spanned by all monomials satisfying three additional properties as follows. First, we require that the monomials have degree at most η in each of the first M' variables $y_1, y_2, \dots, y_{M'}$. Second, their total degree in $y_{M'+1}, y_{M'+2}, \dots, y_M$ is at most μ . Finally, each monomial is nonconstant in at most ν variables taken from $y_{M'+1}, y_{M'+2}, \dots, y_M$. Formally, we adopt the following definition.

DEFINITION 4.13 For $M, M', \eta, \mu, \nu \in \mathbb{N}$ with $M' \leq M$ and $\nu \leq M'' := M - M'$, we set

$$\mathcal{P}_{M', \eta, \mu, \nu} := \text{span}\{\phi_{M, \alpha} : \alpha \in \Sigma_{M', \eta, \mu, \nu}\}, \quad (4.32)$$

where the index set $\Sigma_{M', \eta, \mu, \nu} \subset \mathbb{N}^M$ is given by

$$\Sigma_{M', \eta, \mu, \nu} := \{\alpha = (\alpha', \alpha'') \in \mathbb{N}^{M'} \times \mathbb{N}^{M''} : |\alpha'|_\infty \leq \eta, |\alpha''| \leq \mu, |\text{supp}(\alpha'')| \leq \nu\}. \quad (4.33)$$

In order to prove an approximation property for the polynomial space $\mathcal{P}_{M', \eta, \mu, \nu} \otimes H_0^1(D)$ similar to the one derived in Proposition 4.8 in the context of the h -FE, we first recall that the solution \tilde{u}_M of (3.10) satisfies the estimate (4.6), which we reformulate as follows.

PROPOSITION 4.14 If \tilde{u}_M solves (3.10), then

$$\|\partial_y^\alpha \tilde{u}_M\|_{L^\infty(I^M, H_0^1(D))} \leq c_{a, r, f}^{|\alpha|} |\alpha|! \rho^\alpha \quad \forall \alpha \in \mathbb{N}^M, \quad (4.34)$$

where $\rho^\alpha := \prod_{m=1}^M \rho_m^{\alpha_m}$.

Based on (4.34) we next prove the main approximation result for \tilde{u}_M in the space $\mathcal{P}_{M', \eta, \mu, \nu}$.

PROPOSITION 4.15 If \tilde{u}_M solves (3.10), then there exist $M' \in \mathbb{N}$ and constants $c_1, c_2, c_3 > 0$ depending only on the data a, r, f , such that (recall the notation (4.2))

$$\left\| \tilde{u}_M - \tilde{u}_{M, \Sigma_{M', \eta, \mu, \nu}} \right\|_{L^\infty(I^M, H_0^1(D))} \leq c_1 \left(e^{-c_3 \eta} + e^{c_2 \eta - c_3 \mu} + e^{c_2 \eta - c_3 \nu^{1+\kappa}} \right) \quad (4.35)$$

for any $M, \eta, \mu, \nu \in \mathbb{N}$ with $\nu \leq M'' = M - M'$. Besides,

$$\dim \mathcal{P}_{M', \eta, \mu, \nu} \leq (\eta + 1)^{M'} (M'' + 1)^\nu (\mu + 2)^{\nu+1} (\nu + 1). \quad (4.36)$$

Proof. Let us introduce the notation $y = (y', y'')$ corresponding to the following splitting of the stochastic variable y ,

$$y' = (y_1, y_2, \dots, y_{M'}) \in I^{M'}, \quad y'' = (y_{M'+1}, y_{M'+2}, \dots, y_M) \in I^{M''},$$

where M' will be chosen later. We consider the Taylor expansion of \tilde{u}_M with respect to y around $y = 0$,

$$\tilde{u}_M(y', y'') = \sum_{\substack{\alpha' \in \mathbb{N}^{M'} \\ \alpha'' \in \mathbb{N}^{M''}}} \frac{\partial_{y'}^{\alpha'} \partial_{y''}^{\alpha''} \tilde{u}_M(0)}{\alpha!} y'^{\alpha'} y''^{\alpha''}, \quad (4.37)$$

which converges absolutely for y in a neighbourhood of I^M .

Next, we estimate using Proposition 4.14 the size of that part of the expansion (4.37) which corresponds to the complement of the index set $\Sigma_{\eta, \mu, \nu} \subset \mathbb{N}^M$ given by

$$\Sigma_{\eta, \mu, \nu} := \Sigma'_\eta \times (\Sigma''_\mu \cap \Sigma''_\nu) \quad (4.38)$$

with

$$\begin{aligned}\Sigma'_\eta &:= \{\alpha' \in \mathbb{N}^{M'} : |\alpha'| \leq \eta\} \subset \mathbb{N}^{M'}, \\ \Sigma''_\mu &:= \{\alpha'' \in \mathbb{N}^{M''} : |\alpha''| \leq \mu\} \subset \mathbb{N}^{M''}, \\ \Sigma''_v &:= \{\alpha'' \in \mathbb{N}^{M''} : |\text{supp}(\alpha'')| \leq v\} \subset \mathbb{N}^{M''}.\end{aligned}$$

Note that, due to (4.38),

$$\mathbb{N}^M \setminus \Sigma_{\eta, \mu, v} = ((\mathbb{N}^{M'} \setminus \Sigma'_\eta) \times \mathbb{N}^{M''}) \cup (\Sigma'_\eta \times (\mathbb{N}^{M''} \setminus \Sigma''_\mu)) \cup (\Sigma'_\eta \times (\mathbb{N}^{M''} \setminus \Sigma''_v)). \quad (4.39)$$

Let us denote by T_1, T_2, T_3 those parts of the Taylor expansion (4.37) which correspond to the three disjoint index sets in (4.39), respectively.

An upper bound for T_1 (corresponding to the index set $(\mathbb{N}^{M'} \setminus \Sigma'_\eta) \times \mathbb{N}^{M''}$) follows by a standard p -FE estimate (or using Cauchy formula), due to the analyticity of $\tilde{u}_M(\cdot, y'')$ in a neighbourhood $U_{M'} \times i\mathbb{R}$ of $I^{M'}$ in $\mathbb{C}^{M'}$, uniformly in $y'' \in I^{M''}$,

$$\left\| \sum_{\substack{\alpha' \in \mathbb{N}^{M'} \setminus \Sigma'_\eta \\ \alpha'' \in \mathbb{N}^{M''}}} \frac{\partial_{y'}^{\alpha'} \partial_{y''}^{\alpha''} \tilde{u}_M(0)}{\alpha!} y'^{\alpha'} y''^{\alpha''} \right\|_{L^\infty(I^M, H_0^1(D))} \leq c_{a, f, M'} e^{-c_{1, a, f, M'} \eta}. \quad (4.40)$$

Concerning T_2 (corresponding to the index set $\Sigma'_\eta \times (\mathbb{N}^{M''} \setminus \Sigma''_\mu)$) we have, by (4.34) and the multinomial formula, with $\rho' := (\rho_1, \rho_2, \dots, \rho_{M'})$, $\rho'' := (\rho_{M'+1}, \rho_{M'+2}, \dots, \rho_M)$,

$$\left\| \sum_{\substack{\alpha' \in \Sigma'_\eta \\ \alpha'' \in \mathbb{N}^{M''} \setminus \Sigma''_\mu}} \frac{\partial_{y'}^{\alpha'} \partial_{y''}^{\alpha''} \tilde{u}_M(0)}{\alpha!} y'^{\alpha'} y''^{\alpha''} \right\|_{L^\infty(I^M, H_0^1(D))} \leq \sum_{\substack{\alpha' \in \mathbb{N}^{M'}, \alpha'' \in \mathbb{N}^{M''} \\ |\alpha'| \leq \eta, |\alpha''| \geq \mu+1}} c_{a, f}^{| \alpha |} \frac{| \alpha |!}{\alpha!} \rho^\alpha. \quad (4.41)$$

Using the inequality $|\alpha|! \leq |\alpha'|! \cdot |\alpha''|! \cdot 2^{|\alpha|}$ we separate the variables α', α'' in the summation on the right-hand side of (4.41) and obtain

$$\begin{aligned}\|T_2\|_{L^\infty(I^M, H_0^1(D))} &\leq \sum_{\substack{\alpha' \in \mathbb{N}^{M'} \\ |\alpha'| \leq \eta}} c_{a, f}^{| \alpha' |} \frac{| \alpha' |!}{\alpha'!} \rho'^{\alpha'} \cdot \sum_{l=\mu+1}^{\infty} \sum_{\substack{\alpha'' \in \mathbb{N}^{M''} \\ |\alpha''|=l}} c_{a, f}^l \frac{l!}{\alpha''!} \rho''^{\alpha''} \\ &= \sum_{n=0}^{\eta} (c_{a, f} |\rho'|)^n \cdot \sum_{l=\mu+1}^{\infty} (c_{a, f} |\rho''|)^l \\ &\leq c_{a, r, f}^{\eta+1} (c_{a, f} |\rho''|)^{\mu+1} \leq c_{a, r, f} e^{c_{2, a, r, f} \eta - c_{3, a, f} \mu},\end{aligned} \quad (4.42)$$

where the last two estimates hold if M' is chosen in such a way that

$$c_{a, f} (\rho_{M'+1} + \rho_{M'+2} + \dots) < 1/2.$$

Note that such a choice is always possible due to the decay condition (4.5), and that this is how we determine M' , depending therefore only on the data a, r, f . In turn, the dependence on M' of the two constants in the upper bound (4.40) can be replaced by dependence on r .

Next we estimate T_3 , corresponding to the index set $\Sigma'_\eta \times (\mathbb{N}^{M''} \setminus \Sigma''_\nu)$. We deduce from (4.34) that

$$\begin{aligned} \|T_3\|_{L^\infty(I^M, H_0^1(D))} &\leq \sum_{\substack{\alpha' \in \mathbb{N}^{M'}, \alpha'' \in \mathbb{N}^{M''} \\ |\alpha'| \leq \eta, |\text{supp}(\alpha'')| \geq \nu+1}} c_{a,f}^{|\alpha|} \frac{|\alpha|!}{\alpha!} \rho^\alpha \\ &\leq \sum_{\substack{\alpha' \in \mathbb{N}^{M'} \\ |\alpha'| \leq \eta}} c_{a,f}^{|\alpha'|} \frac{|\alpha'|!}{\alpha'!} \rho^{\alpha'} \cdot \sum_{l=0}^{\infty} \sum_{\substack{\alpha'' \in \mathbb{N}^{M''} \\ |\text{supp}(\alpha'')| \geq \nu+1, |\alpha''|=l}} c_{a,f}^l \frac{l!}{\alpha''!} \rho^{\alpha''}. \end{aligned} \quad (4.43)$$

The first sum (over α') on the right-hand side of (4.43) can be evaluated just as in (4.42), so we only analyse the second one (over l and α''), which we denote in the following by S . To this end, we parameterize the indices α'' through their support (consisting of at least $\nu+1$ integers between 1 and M''), and obtain

$$\begin{aligned} S &= \sum_{l=0}^{\infty} \sum_{j=\nu+1}^{M''} \sum_{1 \leq m_1 < \dots < m_j \leq M''} \sum_{\substack{\alpha'' \in \mathbb{N}^{M''}, |\alpha''|=l \\ \text{supp}(\alpha'') = \{m_1, \dots, m_j\}}} c_{a,f}^l l! \prod_{k=1}^j \frac{\rho_{M'+m_k}^{\alpha''_{m_k}}}{\alpha''_{m_k}!} \\ &= \sum_{l=0}^{\infty} \sum_{j=\nu+1}^{M''} \sum_{1 \leq m_1 < \dots < m_j \leq M''} \sum_{\substack{\alpha'' \in \mathbb{N}^{M''}, |\alpha''|=l \\ \text{supp}(\alpha'') = \{m_1, \dots, m_j\}}} c_{a''} c_{a,f}^l l! \prod_{k=1}^j \frac{\rho_{M'+m_k}^{\alpha''_{m_k}-1}}{(\alpha''_{m_k}-1)!} \end{aligned}$$

with

$$c_{a''} := \prod_{k=1}^j \frac{\rho_{M'+m_k}}{\alpha''_{m_k}} \leq \prod_{k=1}^j \rho_{M'+m_k}. \quad (4.44)$$

Using (4.44) and the multinomial formula, we obtain the estimate

$$\begin{aligned} S &\leq \sum_{l=0}^{\infty} \sum_{j=\nu+1}^{M''} \frac{l!}{(l-j)!} \sum_{1 \leq m_1 < \dots < m_j \leq M''} c_{a,f}^l |\rho''|^{l-j} \prod_{k=1}^j \rho_{M'+m_k} \\ &\leq \sum_{l=0}^{\infty} \sum_{j=\nu+1}^{M''} j! c_{a,f}^j \binom{l}{j} \varepsilon^{l-j} \sum_{1 \leq m_1 < \dots < m_j \leq M''} \rho_{M'+m_1} \cdots \rho_{M'+m_j}, \end{aligned} \quad (4.45)$$

where

$$\varepsilon := c_{a,f}(\rho_{M'+1} + \rho_{M'+2} + \dots + \rho_M).$$

Note that, by increasing M' if necessary (still depending only on the data a, r, f), we can assume, without loss of generality, that $\varepsilon < 1/2$.

Now, from the decay estimate (4.5) we obtain with Lemma A.2 to bound the last sum in (4.45) the bound

$$S \leq \sum_{l=0}^{\infty} \sum_{j=v+1}^{M''} j! c_{a,r,f}^j \binom{l}{j} \varepsilon^{l-j} e^{-c_r j^{1+\kappa}}.$$

Performing first the sum over l via (A.3) in Lemma A.1 and absorbing then the factorial and the exponential functions of j in the last factor, we arrive at ($\varepsilon < 1/2$)

$$S \leq \sum_{j=v+1}^{M''} j! c_{a,r,f}^j \frac{1}{(1-\varepsilon)^{j+1}} e^{-c_r j^{1+\kappa}} \leq c_{a,r,f} e^{-c_{4,a,r,f} v^{1+\kappa}}, \quad (4.46)$$

which ensures via (4.43) that

$$\|T_3\|_{L^\infty(I^M, H_0^1(D))} \leq c_{a,r,f} e^{c_{2,a,r,f} \eta - c_{4,a,r,f} v^{1+\kappa}}. \quad (4.47)$$

Now, (4.35) follows from (4.40), (4.42) and (4.47).

Finally, the dimension estimate (4.36) follows by a counting argument, based on the combinatorial fact that the equation $x_1 + x_2 + \dots + x_q = l$ has exactly $\binom{l}{q}$ solutions $(x_1, x_2, \dots, x_q) \in \mathbb{N}_+^q$, which ensures that

$$\begin{aligned} |\Sigma''_\mu \cap \Sigma''_v| &= \sum_{l=0}^{\mu} \sum_{q=0}^v \binom{M''}{q} \binom{l}{q} \leq (M'' + 1)^v \sum_{l=0}^{\mu} \sum_{q=0}^v \binom{l}{q} \\ &= (M'' + 1)^v \sum_{q=0}^v \binom{\mu + 1}{q + 1} \leq (M'' + 1)^v (\mu + 2)^{v+1} (v + 1), \end{aligned}$$

and the proof is concluded. \square

We recall that we are not asking for an arbitrarily high accuracy in the computation of \tilde{u}_M , since the truncation of the expansion (2.4) of the diffusion coefficient already results in an error between u and u_M of order $O(e^{-c_{1,r} M^\kappa})$ (see Problem 3.5). Making therefore an appropriate choice for the parameters η, μ, v in order to match this accuracy, we arrive at a superalgebraic (though subexponential) convergence rate of the semi-discrete solution of (3.10) with respect to y .

COROLLARY 4.16 If Assumption 2.5 is satisfied and \tilde{u}_M solves (3.10), then there exist $M' \in \mathbb{N}$ and positive constants c_4, c_5, c_6 depending only on a, r, f , such that for

$$\eta := \lceil c_4 M^\kappa \rceil, \quad \mu := \lceil c_5 M^\kappa \rceil, \quad v := \lceil c_6 M^{\kappa/(\kappa+1)} \rceil \quad (4.48)$$

with κ as in Assumption 2.5, we have

$$\|\tilde{u}_M - \tilde{u}_{M, \Sigma_{M, \eta, \mu, v}}\|_{L^\infty(I^M, H_0^1(D))} \leq c_{a,r,f} \exp(-c_{1,r} M^\kappa) \quad (4.49)$$

for all $M \in \mathbb{N}$, $M \geq M_{a,r}$, with

$$N = \dim \mathcal{P}_{M', \eta, \mu, v} \leq \exp(c_{a,r,f} M^{\kappa/(\kappa+1)} \log(M + 2)). \quad (4.50)$$

Proof. We first choose c_4 so that the first term in the upper bound (4.35) matches (4.49). Then we also choose c_5, c_6 (depending on c_2, c_3 in (4.35) and c_4) so that the other two error terms on the right-hand side of (4.35) match (4.49). The dimension estimate (4.50) then follows from (4.36). \square

Combining (4.49) and (4.50), we reformulate the main approximation result of this section as follows.

THEOREM 4.17 If Assumption 2.5 holds, then

$$\inf_{v \in \mathcal{P}_{M', \eta, \mu, v} \otimes H_0^1(D)} \|\tilde{u}_M - v\|_{L^\infty(I^M, H_0^1(D))} \leq c_{1,a,r,f} \exp(-c_{2,a,r,f} (\log N)^{1+\kappa-o(1)}) \quad (4.51)$$

as $M \nearrow \infty$ and for the parameter choice (4.48), where $N = \dim \mathcal{P}_{M', \eta, \mu, v}$ is the number of deterministic diffusion problems to be solved in the physical domain $D \subset \mathbb{R}^d$, and where $\kappa = 1/d$.

Note that the estimated convergence rate (4.51) is asymptotically superalgebraic in the number of deterministic problems N to be solved (due to $\kappa > 0$ in Assumption 2.5), but not asymptotically exponential, as $M \nearrow \infty$ (or, equivalently, as $N \nearrow \infty$).

REMARK 4.18 Our proof of Theorem 4.17 is based on the Taylor expansion of \tilde{u}_M around $y = 0$ (expansion in the standard monomial basis). A similar result can be obtained for the Legendre expansion, by explicitly estimating its coefficients using Proposition 4.4.

4.3 Postprocessing

For brevity of exposition, we only consider here the p -FE-based adapted approximation discussed in Section 4.2.2. Analogous results hold for the h -FE-based chaos approximation of Section 4.2.1.

We show that Theorem 4.17 allows control of the chaos expansion error in the solution to the initial problem (1.2) with respect to a strong (L^∞) topology in the stochastic variable ω .

THEOREM 4.19 Under Assumption 2.5, for

$$u_{M, \Sigma_{M, \eta, \mu, v}}(\cdot, \omega) := \tilde{u}_{M, \Sigma_{M, \eta, \mu, v}}(\cdot, X_1(\omega), \dots, X_M(\omega)) \in H_0^1(D) \quad P\text{-a.e. } \omega \in \Omega,$$

we have, with κ as in Assumption 2.5,

$$\|u - u_{M, \Sigma_{M, \eta, \mu, v}}\|_{L^\infty(\Omega, H_0^1(D))} \leq c_{1,a,r,f} \exp(-c_{2,a,r,f} (\log N)^{1+\kappa-o(1)}) \quad (4.52)$$

as $M \nearrow \infty$ and for the choice (4.48) of discretization parameters.

Proof. The claim follows immediately from (4.51) (adapted chaos error estimate) and (3.5) (fluctuation truncation error estimate), taking into account the relationship between the number M of stochastic variables y_m and the number N of deterministic problems to be solved given by (4.50). \square

REMARK 4.20 The boundedness of the k th moment operator between $L^\infty(\Omega, H_0^1(D))$ and $\underbrace{H_0^1(D) \otimes \dots \otimes H_0^1(D)}_{k \text{ times}}$ ensures an upper bound similar to (4.52) also for the y -semidiscretization error

in these moments.

5. Sparse collocation approximation

The construction of the p -FE-based adapted chaos expansion in Section 4.2.2 shows that the sG method can achieve convergence rates which are superalgebraic in terms of the number N of deterministic problems to be solved. This should be compared with the rate $N^{-1/2}$ exhibited by MC approximations.

This comparison which appears unfavourable for MC methods must be amended by the fact that the sG discretization requires the solution of a very large, linear system of equations (its matrix being the Kronecker Product matrix of the *stochastic mass matrix* from the Galerkin projection in the y variable and the stiffness matrix of the deterministic problem in the physical domain).

We therefore propose and discuss in this section another (Chebyshev type) interpolant of \tilde{u}_M , for which superalgebraic convergence similar to that of the p -FE-based adapted chaos expansion shown in Section 4.2.2) holds, but whose computation only requires sampling of \tilde{u}_M at prescribed values of $y \in I^M$; it can, in particular, be obtained with the same computational work as a MC approximation based on N samples, which are, like MC approximations, computable independently and in parallel.

Throughout this section, we focus only on the p -FE-based adapted chaos expansion which, compared to its h -FE counterpart in Section 4.2.1, best exploits the analyticity of the solution with respect to the stochastic variable y .

We start by observing that the definition of the adapted sparse polynomial space in Definition 4.13 together with the parameter choice (4.48) show that \tilde{u}_M , which depends in principle on M variables y_1, \dots, y_M , can be approximated to the desired accuracy (see (4.49)) by a linear combination of functions of at most $\nu := \lceil c_6 M^{\kappa/(\kappa+1)} \rceil \ll M$ (as $M \nearrow \infty$) variables taken from $\{y_1, \dots, y_M\}$, where we assume without loss of generality that $M' = 0$. We therefore construct the interpolant of \tilde{u}_M by taking Chebyshev-type interpolants of all functions in the representation of \tilde{u}_M .

We start by introducing several notations. If $I = [-1/2, 1/2]$, B is a Banach space and $g: I^M \rightarrow B$ a B -valued analytic function, we denote for every subset S of $\{1, 2, \dots, M\}$ by $g_S: I^{|S|} \rightarrow B$ the restriction of g to $y_m = 0$ for all $m \notin S$. Note that by definition we then have $g_\emptyset = g(0)$. With these notations the following combinatorial identity holds.

PROPOSITION 5.1 If $g: I^M \rightarrow B$ is analytic and $\nu \in \mathbb{N}$, $\nu \leq M$, then

$$\sum_{\substack{S \subset \{1, \dots, M\} \\ |S| \leq \nu}} (-1)^{\nu - |S|} \binom{M - |S|}{\nu - |S|} g_S(y) = \sum_{\substack{\alpha \in \mathbb{N}^M \\ |\text{supp}(\alpha)| = \nu}} \frac{\partial^\alpha g(0)}{\alpha!} y^\alpha \quad \forall y \in I^M. \quad (5.1)$$

Proof. The formula is a consequence of the inclusion–exclusion principle. Denoting for $\alpha \in \mathbb{N}^M$ by $t_\alpha := \partial^\alpha g(0) y^\alpha / \alpha!$ the general term of the Taylor expansion of g around 0, we expand all functions g_S appearing in (5.1) in Taylor series around 0,

$$g_S(y) = \sum_{\substack{\beta \in \mathbb{N}^M \\ \text{supp}(\beta) \subseteq S}} t_\beta,$$

and calculate the resulting coefficient of t_α on the left-hand side of (5.1) for arbitrary $\alpha \in \mathbb{N}^M$ with $|\text{supp}(\alpha)| \leq \nu$. Note first that for $|\text{supp}(\alpha)| = \nu$, the term t_α only appears in the Taylor expansion of h_S with $S = \text{supp}(\alpha)$, so that its coefficient on the left-hand side of (5.1) equals 1. For $r := |\text{supp}(\alpha)| < \nu$, the coefficient of t_α equals then ($s := |S|$)

$$\begin{aligned} \sum_{s=0}^{\nu} (-1)^{\nu-s} \binom{M-s}{\nu-s} \binom{M-r}{s-r} &= \sum_{s=r}^{\nu} (-1)^{\nu-s} \binom{\nu-r}{\nu-s} \binom{M-r}{\nu-r} \\ &= (1-1)^{\nu-r} \binom{M-r}{\nu-r} = 0, \end{aligned}$$

which concludes the proof. \square

Proposition 5.1 yields the following representation of g .

COROLLARY 5.2 If $g: I^M \rightarrow B$ is analytic and $v \in \mathbb{N}$, $v \leq M$, then

$$g(y) - \sum_{\substack{S \subset \{1, \dots, M\} \\ |S| \leq v}} \gamma_{M,|S|,v} g_S(y) = \sum_{\substack{\alpha \in \mathbb{N}^M \\ |\text{supp}(\alpha)| > v}} \frac{\partial^\alpha g(0)}{\alpha!} y^\alpha \quad \forall y \in I^M, \quad (5.2)$$

where

$$\gamma_{M,s,v} = \sum_{r=s}^v (-1)^{r-s} \binom{M-s}{r-s}. \quad (5.3)$$

Proof. Formula (5.2) follows easily from (5.1) after replacing v by r and taking the sum over r from 0 to v . \square

We note for later use that the coefficients γ in the sparse interpolation operator in (5.2) which are given by (5.3) satisfy the following estimate.

LEMMA 5.3 If $v \in \mathbb{N}$, for $\gamma_{M,s,v}$ given by (5.3) we have

$$\sum_{\substack{S \subset \{1, \dots, M\} \\ |S| \leq v}} |\gamma_{M,|S|,v}| \leq (v+1)(M+1)^{2v}. \quad (5.4)$$

Proof. The formula follows using the inequality $\binom{n}{k} \leq (n+1)^k$ to bound (5.3) from above and noting that the number of terms in the summation on the left-hand side of (5.4) does not exceed $(v+1)(M+1)^v$. \square

Formula (5.2) allows control of the error in the approximation of g as a function of $y \in I^M$ by an explicit linear combination of restrictions of g to lower-dimensional hyperplane sections of I^M . We will apply it to approximate $\tilde{u}_{M,h}: I^M \rightarrow V_h$, the semi-discrete (in the deterministic variable $x \in D$) solution to (3.10). More precisely, $V_h \subset H_0^1(D)$ denotes here a FE space in the physical domain D (h stands for the discretization parameter), so that $\tilde{u}_{M,h}: I^M \rightarrow V_h$ solves the deterministic, parametric elliptic problem (in variational form)

$$-\text{div}(\tilde{a}_M(\cdot, y) \nabla \tilde{u}_{M,h}(\cdot, y)) = f(\cdot) \quad \text{in } V_h^* \quad \forall y \in I^M. \quad (5.5)$$

We first note that uniform ellipticity of (3.10) for $M \geq M_{a,r}$ ensures the quasioptimality of the FE solution $\tilde{u}_{M,h}$, uniformly in $y \in I^M$,

$$\|\tilde{u}_M(\cdot, y) - \tilde{u}_{M,h}(\cdot, y)\|_{H_0^1(D)} \leq c_a \inf_{v \in H_0^1(D)} \|\tilde{u}_M(\cdot, y) - v\|_{H_0^1(D)} \quad \forall y \in I^M, \quad (5.6)$$

which in turn can be combined with the approximation property of V_h to derive uniform in y , semi-discrete (in the deterministic variable $x \in D$) convergence rates for $\tilde{u}_{M,h}(\cdot, y)$, under standard data-regularity assumptions.

We apply Corollary 5.2 with $g := \tilde{u}_{M,h}$, whose analyticity domain in y is, just as in the case of \tilde{u}_M , the complex neighbourhood $U^M \times i\mathbb{R}$ of I , with $I \subset U \subset \mathbb{R}$. The right-hand side of (5.2) then satisfies an estimate identical to (4.47) with $M' = 0$, so that we obtain

PROPOSITION 5.4 For $v \in \mathbb{N}$, $v \leq M$, we have

$$\left\| \tilde{u}_{M,h} - \sum_{\substack{S \subset \{1, \dots, M\} \\ |S| \leq v}} \gamma_{M,|S|,v} \tilde{u}_{M,h,S} \right\|_{L^\infty(I^M, H_0^1(D))} \leq c_{a,r,f} e^{-c_{4,a,r,f} v^{1+\kappa}}, \quad (5.7)$$

where the coefficients γ are given by (5.3).

Note that, in order to match the required accuracy $O(\exp(-c_{1,r} M^\kappa))$ in the computation of $\tilde{u}_{M,h}$, we will later choose v as in (4.48).

The approximation of $\tilde{u}_{M,h}$ considered so far is only semi-discrete (in the deterministic variable $x \in D$), and full discretization is needed in order to allow computation of $\tilde{u}_{M,h}$. We consider therefore for a fixed order $\mu \in \mathbb{N}$ (to be chosen later), the tensor-product Chebyshev interpolants $\mathcal{J}_\mu \tilde{u}_{M,h,S}$ of $\tilde{u}_{M,h,S}$ for all S as in (5.7), i.e.

$$\mathcal{J}_{v,\mu} \tilde{u}_{M,h} := \sum_{\substack{S \subset \{1, \dots, M\} \\ |S| \leq v}} \gamma_{M,|S|,v} \mathcal{J}_\mu \tilde{u}_{M,h,S}. \quad (5.8)$$

We next carry out the error analysis for the interpolant (5.8). We start by recalling that for the tensor-product, Chebyshev interpolant of analytic functions holds the following pointwise estimate of the interpolation error.

THEOREM 5.5 Suppose B is a Banach space and $g: [-1, 1]^M \rightarrow B$ has an analytic extension to $\mathcal{E}_\theta := \mathcal{E}_{\theta_1} \times \mathcal{E}_{\theta_2} \times \dots \times \mathcal{E}_{\theta_M} \subset \mathbb{C}^M$, where for $\theta > 1$, \mathcal{E}_θ denotes the closed ellipse with foci at ± 1 and sum of its semi-axes equal to θ . Then for any $\underline{\mu} = (\mu_1, \dots, \mu_M) \in \mathbb{N}^M$, the anisotropic Chebyshev interpolant $\mathcal{J}_{\underline{\mu}} g$ of g satisfies

$$\|g - \mathcal{J}_{\underline{\mu}} g\|_{L^\infty(I^M, B)} \leq 2^{M/2+1} \prod_{m=1}^M \frac{1}{\sqrt{1 - \theta_m^{-2}}} \left(\sum_{m=1}^M \theta_m^{-2\mu_m} \right)^{1/2} \|g\|_{L^\infty(\mathcal{E}_{\underline{\theta}}, B)}. \quad (5.9)$$

After applying a scaling argument on $I = [-1/2, 1/2]$, we apply Theorem 5.5 with $g = \tilde{u}_{M,h,S}$. From Proposition 5.4 and Theorem 5.5, we obtain the following error estimate of the interpolant given by (5.8).

PROPOSITION 5.6 Under Assumption 2.5, for the interpolant (5.8) of $\tilde{u}_{M,h}$, the solution to (5.5), we have that

$$\|\tilde{u}_{M,h} - \mathcal{J}_{v,\mu} \tilde{u}_{M,h}\|_{L^\infty(I^M, H_0^1(D))} \leq c_{a,r,f} \left(e^{-c_{4,a,r,f} v^{1+\kappa}} + (M+1)^{2v} e^{c_{5,a,r} v - c_{6,a,r} \mu} \right). \quad (5.10)$$

Moreover, the construction of $\mathcal{J}_{v,\mu} \tilde{u}_{M,h}$ only requires sampling of $\tilde{u}_{M,h}$ at not more than

$$N := (v+1)(M+1)^v(\mu+1)^v \quad (5.11)$$

sampling points $y \in I^M$ using any solver of the deterministic diffusion problem with FE space V_h .

Proof. Using the triangle inequality and (5.4), (5.7), we have in $L^\infty(I^M, H_0^1(D))$,

$$\begin{aligned} & \|\tilde{u}_{M,h} - \mathcal{J}_{v,\mu} \tilde{u}_{M,h}\| \\ & \leq \left\| \tilde{u}_{M,h} - \sum_{\substack{S \subset \{1,\dots,M\} \\ |S| \leq v}} \gamma_{M,|S|,v} \tilde{u}_{M,h,S} \right\| + \sum_{\substack{S \subset \{1,\dots,M\} \\ |S| \leq v}} |\gamma_{M,|S|,v}| \|\tilde{u}_{M,h,S} - \mathcal{J}_\mu \tilde{u}_{M,h,S}\| \\ & \leq c_{a,r,f} e^{-c_{4,a,r,f} v^{1+\kappa}} + (v+1)(M+1)^{2v} \cdot \max_{\substack{S \subset \{1,2,\dots,M\} \\ |S| \leq \mu}} \|\tilde{u}_{M,h,S} - \mathcal{J}_\mu \tilde{u}_{M,h,S}\|. \end{aligned} \quad (5.12)$$

In order to estimate the interpolation error $\|\tilde{u}_{M,h,S} - \mathcal{J}_\mu \tilde{u}_{M,h,S}\|$ via Theorem 5.5, we first choose $\theta > 1$ (depending on the fluctuation r and the bounds on a) such that $\tilde{u}_{M,h}$ is analytic in \mathcal{E}_θ^M for all $M \in \mathbb{N}_+$. Note that this choice is possible due to (4.5). We apply now Theorem 5.5 with $g = \tilde{u}_{M,h,S}$ (and M replaced by $|S| \leq v$), noting that

$$\|\tilde{u}_{M,h,S}\|_{L^\infty(\mathcal{E}_\theta^M, H_0^1(D))} \leq c_{a,r,f}$$

for any $S \subset \{1, 2, \dots, M\}$ and any $M \in \mathbb{N}_+$, $M \geq M_{a,r}$ due to the uniform in y ellipticity of (5.5). We obtain

$$\|\tilde{u}_{M,h,S} - \mathcal{J}_\mu \tilde{u}_{M,h,S}\| \leq c_{a,r,f} e^{c_{5,a,r} v - c_{6,a,r} \mu} \quad \forall S \subset \{1, 2, \dots, M\}, \quad \forall M \in \mathbb{N}_+, \quad M \geq M_{a,r},$$

which we now combine with (5.12) to obtain (5.10).

The estimated computational cost for the collocation approximation (5.8) of solving N elliptic problems with N as in (5.11) of (5.8) follows by noting that $\mathcal{J}_\mu \tilde{u}_{M,h,S}$ can be obtained by sampling of $\tilde{u}_{M,h,S}$ at $(\mu+1)^{|S|}$ (Chebyshev) points in the coordinates y_m with $m \in S$, plus a counting argument similar to the one used to derive (4.36). \square

An appropriate choice of the parameters v, μ ensures the desired accuracy of the interpolant (5.8).

THEOREM 5.7 If Assumption 2.5 is satisfied and $\tilde{u}_{M,h}$ solves (5.5), then there exist positive constants c_7, c_8 and $M_{a,r}$ depending only on a, r, f , such that for

$$v := \lceil c_7 M^{\kappa/(\kappa+1)} \rceil, \quad \mu := \lceil c_8 M^\kappa \rceil, \quad (5.13)$$

we have

$$\|\tilde{u}_{M,h} - \mathcal{J}_{v,\mu} \tilde{u}_{M,h}\|_{L^\infty(I^M, H_0^1(D))} \leq c_{a,r,f} \exp(-c_{1,r} M^\kappa) \quad (5.14)$$

for all $M \in \mathbb{N}$, $M \geq M_{a,r}$.

Moreover, the construction of $\mathcal{J}_{v,\mu} \tilde{u}_{M,h}$ only requires sampling of $\tilde{u}_{M,h}$ in at most

$$N \leq \exp(c_{a,r,f} M^{\kappa/(\kappa+1)} \log(M+1))$$

points $y \in I^M$, each sample corresponding to the solution of the deterministic diffusion problem with FE space V_h .

Proof. We first choose c_7 and then c_8 such that the two estimated error contributions on the right-hand side of (5.10) match the right-hand side of (5.14). The estimated computational cost for the interpolant (5.8) then follows using (5.13) in (5.11). \square

Acknowledgements

This work was supported in part under the network *Breaking Complexity* of the European commission (contract number HPRN-CT-2002-00286) with support by the Swiss Federal Office for Science and Education under grant No. BBW 02.0418.

REFERENCES

- BABUŠKA, I., TEMPONE, R. & ZOURARIS, G. E. (2004) Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.*, **42**, 800–825 (electronic).
- DAVIS, P. J. (1963) *Interpolation and Approximation*. New York: Blaisdell Publishing Co.
- DETTINGER, M. D. & WILSON, J. L. (1981) First order analysis of uncertainty in numerical models of ground-water flow, 1, mathematical development. *Water Resour. Res.*, **17**, 149–161.
- FRAUENFELDER, P., SCHWAB, C. & TODOR, R. A. (2005) Finite elements for elliptic problems with stochastic coefficients. *Comput. Methods Appl. Mech. Eng.*, **194**, 205–228.
- GHANEM, R. G. & SPANOS, P. D. (1997) Spectral techniques for stochastic finite elements. *Arch. Comput. Methods Eng.*, **4**, 63–100.
- KELLER, J. B. (1964) Stochastic equations and wave propagation in random media. *Proc. Sympos. Appl. Math.*, vol. XVI. (R. Bellmann ed.). Providence, RI: American Mathematical Society, pp. 145–170.
- KLEIBER, M. & HIEN, T. D. (1992) *The Stochastic Finite Element Method*. Chichester, UK: John Wiley.
- KÖNIG, H. (1986) *Eigenvalue Distribution of Compact Operators. Operator Theory: Advances and Applications*, vol. 16. Basel, Switzerland: Birkhäuser.
- LOÈVE, M. (1977) *Probability Theory. I*. 4th edn. *Graduate Texts in Mathematics*, vol. 45. New York: Springer.
- LOÈVE, M. (1978) *Probability Theory. II*. 4th edn. *Graduate Texts in Mathematics*, vol. 46. New York: Springer.
- MATTHIES, H. G. & KEESE, A. (2005) Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Comput. Methods Appl. Mech. Eng.*, **194**, 1295–1331.
- PIETSCH, A. (1987) *Eigenvalues and s-Numbers. Cambridge Studies in Advanced Mathematics*, vol. 13. Cambridge: Cambridge University Press.
- PINKUS, A. (1985) *n-Widths in Approximation Theory*. New York: Springer.
- SCHOUTENS, W. (2000) *Stochastic Processes and Orthogonal Polynomials*. Lecture Notes in Statistics, vol. 146. New York: Springer.
- SCHWAB, C. & TODOR, R. A. (2006) Karhunen–Loève approximation of random fields in domains by generalized fast multipole methods. *Journ. Comp. Phys.*, **217**, 100–122.
- TODOR, R. A. (2006) Robust eigenvalue computation for smoothing operators. *SIAM J. Numer. Anal.* (to appear).
- TODOR, R. A. (2005) Sparse perturbation algorithms for elliptic PDE's with stochastic data. *Ph.D. Thesis No. 16192*, ETH Zurich, Switzerland.
- WIENER, N. (1938) The homogeneous chaos. *Am. J. Math.*, **60**, 897–936.
- XIU, D. & KARNIADAKIS, G. E. (2002) The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, **24**, 619–644 (electronic).

Appendix A

Here, we state and prove two technical lemmas which we used repeatedly in the derivation of convergence and complexity bounds.

LEMMA A.1 For any $t \in [0, 1)$ and $j, L \in \mathbb{N}$ with $j \leq L$, we have

$$\sum_{n=0}^{\infty} \binom{L+n}{j} t^n \leq (L+1)^j (1-t)^{-j-1}. \quad (\text{A.1})$$

Proof. Using the factorial representation of the binomial coefficients it is easy to see that

$$\binom{L+n}{j} \leq (L+1)^j \binom{j+n}{j} \quad \forall n \in \mathbb{N},$$

which ensures

$$\sum_{n=0}^{\infty} \binom{L+n}{j} t^n \leq (L+1)^j \sum_{n=0}^{\infty} \binom{j+n}{j} t^n. \quad (\text{A.2})$$

Denoting by S_j the sum on the right-hand side of (A.2), the binomial identity

$$\binom{j+n}{j} = \binom{j+n-1}{j} + \binom{j+n-1}{j-1}$$

leads to the recursive formula $S_j = tS_j + S_{j-1}$, which shows that $(S_0 = (1-t)^{-1})$

$$S_j = (1-t)^{-j-1}. \quad (\text{A.3})$$

(A.1) then follows from (A.2) and (A.3). \square

We now prove that, if $y, \kappa > 0$ and $j \in \mathbb{N}_+$, the sum of the series with general term $\exp(-y \sum_{i=1}^j m_i^\kappa)$ indexed over $1 \leq m_1 < \dots < m_j < \infty$ is, qualitatively and uniformly in $j \in \mathbb{N}_+$, just as large as the leading term, corresponding to $m_i = i$ for all $1 \leq i \leq j$. More precisely, we have the following result.

LEMMA A.2 If $\kappa > 0$ and $x > y > z > 0$, then there exist $c_{\kappa,x,y}, c_{\kappa,y,z} > 0$ such that

$$c_{\kappa,x,y} \exp\left(-x \frac{1}{1+\kappa} j^{1+\kappa}\right) \leq \sum_{1 \leq m_1 < \dots < m_j < \infty} \prod_{i=1}^j \exp(-y m_i^\kappa) \leq c_{\kappa,y,z} \exp\left(-z \frac{1}{1+\kappa} j^{1+\kappa}\right) \quad (\text{A.4})$$

for all $j \in \mathbb{N}_+$.

Proof. For $y > 0$ and $j \in \mathbb{N}_+$, $J \in \mathbb{N}_+ \cup \{\infty\}$ with $j \leq J$, we set

$$S_{j,J} := \sum_{1 \leq m_1 < \dots < m_j \leq J} \prod_{i=1}^j \exp(-y m_i^\kappa). \quad (\text{A.5})$$

The lower bound in (A.4) follows by observing that the sum in (A.5) contains the term corresponding to $m_i = i$ for all $1 \leq i \leq j$, so that

$$S_{j,J} \geq \exp\left(-y \sum_{i=1}^j i^\kappa\right),$$

where

$$\sum_{i=1}^j i^\kappa \leq (j+1)^{1+\kappa} \int_0^1 x^\kappa dx = \frac{1}{1+\kappa} (j+1)^{1+\kappa}.$$

It remains to prove the upper bound of the sum in (A.4). It follows from (A.5) that the sequence $(S_{j,j})_{j \in \mathbb{N}_+}$ is rapidly decaying, i.e.

$$S_{j,j} \leq c_{\kappa,y,\beta} \beta^j \quad \forall j \in \mathbb{N}_+, \quad \forall \beta > 0. \quad (\text{A.6})$$

We also derive from (A.5) the recursive formula

$$\begin{aligned} S_{j,J+1} &= S_{j,J} + \exp(-y(J+1)^\kappa) \sum_{1 \leq m_1 < \dots < m_{j-1} \leq J} \prod_{k=1}^{j-1} \exp(-y m_k^\kappa) \\ &= S_{j,J} + \exp(-y(J+1)^\kappa) S_{j-1,J}. \end{aligned} \quad (\text{A.7})$$

By induction on j in (A.7) we immediately see that

$$S_{j,J} < S_{j,\infty} = \lim_{J \nearrow \infty} S_{j,J} < \infty \quad \forall j \in \mathbb{N}_+,$$

and that

$$S_{j,\infty} \leq S_{j,j} + \sum_{i=j+1}^{\infty} \exp(-y i^\kappa) \cdot S_{j-1,\infty}. \quad (\text{A.8})$$

Now, for an arbitrary $\gamma \in]0, 1[$ we have, for j large enough ($j \geq j_0$, with j_0 depending on y, κ, γ),

$$\sum_{i=j+1}^{\infty} \exp(-y i^\kappa) \leq \gamma,$$

which ensures via (A.8)

$$S_{j,\infty} \leq S_{j,j} + \gamma S_{j-1,\infty} \quad \forall j \geq j_0. \quad (\text{A.9})$$

We deduce from (A.6) and (A.9) that

$$S_{j,\infty} \leq c_{\kappa,y,\beta} (\gamma + \beta)^j + \gamma^{j-j_0+1} S_{j_0-1,\infty},$$

which shows that $S_{j,\infty} \rightarrow 0$ as $j \nearrow \infty$, by choosing β such that $\gamma + \beta < 1$. The sequence $(S_{j,\infty})_{j \in \mathbb{N}_+}$ is in particular bounded, i.e.

$$S_{j,\infty} \leq c_{\kappa,y} \quad \forall j \in \mathbb{N}_+. \quad (\text{A.10})$$

Since this inequality holds for any $y > 0$, the conclusion then follows from (A.10) upon replacing y by $y - z$ and noting that

$$\sum_{i=1}^j i^\kappa \geq j^{1+\kappa} \int_0^1 x^\kappa dx = \frac{1}{1+\kappa} j^{1+\kappa}.$$

□